



HAL
open science

Building and comparing values from different sources

Basile Garcia

► **To cite this version:**

Basile Garcia. Building and comparing values from different sources. Neuroscience. Université Paris sciences et lettres, 2022. English. NNT : 2022UPSLE016 . tel-04461363

HAL Id: tel-04461363

<https://theses.hal.science/tel-04461363v1>

Submitted on 16 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL
Préparée à l'École Normale Supérieure

Building and comparing values from different sources

Soutenue par

Basile GARCIA

Le 04/03/2022

École doctorale n°158

**Cerveau, cognition,
comportement**

Spécialité

**Neurosciences
computationnelles**

Composition du jury :

| | |
|--|---------------------------|
| Wim DE NEYS Sorbonne University | <i>Président</i> |
| Giorgia ROMAGNOLI University of Amsterdam | <i>Rapporteur</i> |
| Valérie DUFOUR Strasbourg University | <i>Rapporteur</i> |
| Ralph HERTWIG Max Planck Institute for Human Development | <i>Examineur</i> |
| Stefano PALMINTERI ENS - PSL Research University | <i>Directeur de thèse</i> |
| Sacha BOURGEOIS-GIRONDE ENS - PSL Research University | <i>Directeur de thèse</i> |

Abstract

Subjective value is an ubiquitous construct in the study of decision-making. In this literature, individuals' decisions are often conceived as a two-step process. They first assign values to the available options, and then choose the option with the highest value. Typically, the explanatory variable "value" thus quantifies the intensity of a preference for one option over others. This conceptualization stems from the intersection of several disciplines, notably (behavioral) economics as well as experimental psychology and neuroscience. In retrospect, the construction of subjective value, and its purpose in decision-making, can be traced back to two historic experimental approaches. The *description paradigm* provides the decision-maker with full knowledge of options and associated consequences. This paradigm is anchored to economic assumptions of rationality and is historically related to the quantification of economic value, with ramifications in moral philosophy (utilitarianism). By contrast, the *experience paradigm* is rooted in animal reinforcement learning study, where the lack of information implies to learn by trial-and-error. Both paradigms concurrently developed methods to elicit subjective values. Interestingly, this resulted in behavioral discrepancies, known as the *description-experience gap*.

About thirty years ago, the field of value-based decision-making emerged when scholars with different backgrounds attempted to merge both fields and provided a neurobiological ground for the concept of subjective value. The common currency hypothesis posits that items which fundamentally differ in nature (say water and a car) can be compared through the mapping of each item's attributes on a common scale, forming a subjective value associated to each item. This scaling (and comparison) process is thought to be implemented in the brain, and neurally represented by the firing activity of dopaminergic neurons. However, attributing this activity to value *per se* is difficult. Value as a predictive variable often correlates with attention, arousal or salience. In addition, assuming values acquired from learned experiences are encoded on a neural common scale, it remains unclear to what extent those values are transformed during both the coding and the retrieval process. A strong version of the common currency hypothesis would postulate that they are sufficiently maintained to be properly compared to other kinds of values, such as the ones built via symbolic descriptions of decision variables.

Through a review of the literature, we asked whether the observed gap between description- and experience-based choices might hinder our ability to build mechanistic models of decision. Thereafter, in the main experimental study presented in this work, we questioned the possibility of directly comparing experiential subjective values to external symbolic ones represented in the environment, thus aimed to behaviorally test some of common currency predictions. More generally, we challenged the traditional two-step model of value based decision-making in humans, which posits that individuals go through a valuation stage and a choice stage, to ultimately maximize expected value.

Remerciements

Il est notoire que la préparation et rédaction d'une thèse de doctorat est un travail éprouvant. Cependant, la trajectoire qui amène à produire une thèse, loin d'être strictement individuelle, est emmaillée de rencontres, qui l'on fait advenir. Cette dimension collective de la *science normale*, c'est ce qui m'a permis d'arriver au présent document, en esquivant les conséquences néfastes de l'activité scientifique. En effet, telle que ses institutions l'organisent, cette dernière affecte encore trop durement beaucoup de doctorant-e-s (Woolston, 2019; Satinsky et al., 2021). Ainsi, j'aimerais ici remercier toutes les personnes ¹ qui, par leur présence, leur soutien, leurs paroles, leurs suggestions, ont contribué à rendre possible l'aboutissement de cette thèse.

J'aimerais remercier en premier lieu mes directeurs, Stefano et Sacha. Cette opportunité que vous m'avez offerte m'a arraché à une vie de berger en Nouvelle-Zélande, on espère que c'est pour le mieux ! Je ne pourrais énumérer les qualités variées qui vous ont caractérisé pendant ces trois ans, heureusement, l'exercice des remerciements me contraint à rester succinct. Stefano, ta disponibilité quasi-permanente, ton implication professionnelle et humaine auprès de tes doctorants, ainsi que l'efficacité de tes analyses scientifiques, sont autant de qualités qui t'honorent en tant qu'encadrant. Sacha, ta capacité à prendre de la distance par rapport aux objets scientifiques et la profusion d'idées et d'intuitions qui en découlent, ont contribué à rendre nos discussions à chaque fois plus inspirantes. Ces moments éparses ont pu agir comme des bouffées d'air lorsque j'étais noyé dans des problèmes technico-techniques.

Je remercie également mon jury, en particulier Giorgia Romagnoli et Valérie Dufour qui ont accepté le travail fastidieux de lire et évaluer ma thèse (et dont les retours pertinents et positifs m'ont ravi !). Aussi, merci à Wim De Neys pour avoir accepté de présider le jury, et Ralph Hertwig pour sa présence et dont les travaux ont inspiré les études incluses dans cette thèse.

Je souhaite remercier tout particulièrement l'équipe HRL, dont les membres animent (ou ont animé) mon quotidien : Anis, Germain, Sabrine, Ali, Stefano, Antonis (pour les divers moments agréables et discussions riches autour de sujets variés, allant du rap au machine learning, en passant par l'histoire politique Iranienne ou Grecque). En outre, merci à Zoë (hâte de se faire un petit concert), Hernan (réfèrent Péronisme et BOTW) et Magda (référente en post-soviétisme, impressionné dernièrement par ta tenacité en fin de soirée). Un grand merci aussi à la team (ou fratrie) "discord" : Sophie (ma soeur de thèse), Fabien (cousin de thèse pourrait-on dire), et Henri (petit-frère de thèse). Je souhaite que nos (d)ébats scientifiques, politiques et vidéoludiques s'éternisent.

¹Je m'excuse par avance auprès de celles et ceux que je vais oublier...

Je remercie également les différents membres du LNC² (et personnes gravitant autour), sans qui ces trois dernières années auraient été bien moins plaisantes : Clémence (dont la volonté de se conformer aux traditionnels et désuets saluts matinaux est sans limite), Julie (nous partageons la peur de la menace wokiste), Jun ("tu connais le cinéma de Na Hong-jin ?"), Margaux (dont la variété des engagements scientifiques, allant de la séquestration de participants aux rixes avec des lièvres variables, m'impressionne), et Marine (sans qui le labo s'effondrerait). Merci également à Damiano (bientôt une carrière chez McKinsey ?), Aurélien W. (dont les excellents spectacles et autres soirées à domiciles me resteront en mémoire), Victor (au plaisir de discuter sciences sociales ?), Rocco (pour les conseils avisés), Camille (bon courage pour la gestion du serveur ! Et regarde "Scrubs" !) Morgan (pour les débuts de discussions politiques, malheureusement souvent écourtées), ou encore Maël et Valentin (qui sont bien inspirés scientifiquement mais moins lorsqu'il s'agit de chambrer).

Je n'oublie évidemment pas mes proches, famille et amis, pour qui j'ai une gratitude et une affection infinie, et qui ont pu m'assister et me soutenir durant les différentes étapes de mon parcours personnel et professionnel. Je souhaite, et j'espère, leur être d'une aide équivalente.

Un merci particulier à Aurélien N., dont la formation intellectuelle, et par la suite l'amitié, ont été déterminants dans mon parcours universitaire.

Enfin, Juliette, sans qui je n'aurais tout simplement pas tenu...

Contents

| | | |
|-------|---|----|
| 0 | INTRODUCTION | 1 |
| 1 | UTILITY THEORIES AND DECISION-MAKING IN ECONOMICS | 5 |
| 1.1 | Toward a subjective theory of value | 5 |
| 1.1.1 | An objective theory of value: Labor theory of value | 5 |
| 1.1.2 | A subjective theory of value: utility | 7 |
| 1.1.3 | The expected utility hypothesis | 9 |
| 1.1.4 | Decision under risk | 10 |
| 1.1.5 | Moral utilitarianism and hedonistic utility | 11 |
| 1.1.6 | Marginal Revolution | 12 |
| 1.1.7 | Ordinal Revolution | 13 |
| 1.2 | Axiomatic utility | 14 |
| 1.2.1 | Revealed preferences | 14 |
| 1.2.2 | Risk-attitudes | 15 |
| 1.2.3 | Von Neumann-Morgenstern utility theorem | 16 |
| 1.2.4 | Allais paradox | 19 |
| 1.3 | Behavioral models of value and decision-making | 21 |
| 1.3.1 | Anomalies in decision-making | 21 |
| 1.3.2 | Judgment and Decision-Making | 22 |
| 1.3.3 | Different research programs | 23 |
| 1.3.4 | Value models | 24 |
| 1.3.5 | Comparison-based and value-free models | 29 |
| 1.4 | Summary | 38 |
| 2 | REINFORCEMENT LEARNING | 41 |
| 2.1 | Behavioral reinforcement learning | 41 |
| 2.1.1 | Classical conditioning | 41 |
| 2.1.2 | Operant conditioning | 42 |
| 2.1.3 | Behaviorism | 43 |
| 2.1.4 | Blocking effect | 46 |
| 2.2 | Computational reinforcement learning | 47 |
| 2.2.1 | Basic principles | 47 |
| 2.2.2 | Prediction Problem and Control Problem | 49 |
| 2.2.3 | Rescorla-Wagner Model | 50 |
| 2.2.4 | Delta rule for Neural Nets | 51 |

| | | |
|----------|--|------------|
| 2.2.5 | Temporal difference learning | 52 |
| 2.2.6 | Actor-critic | 55 |
| 2.2.7 | Q-Learning | 57 |
| 2.2.8 | Action selection and the exploration-exploitation trade-off | 58 |
| 2.2.9 | Value-free models | 59 |
| 2.3 | Neural reinforcement learning | 62 |
| 2.3.1 | Value-based decision-making | 62 |
| 2.3.2 | The prediction-error in monkeys | 62 |
| 2.3.3 | A neural common currency in humans | 64 |
| 2.4 | Summary | 67 |
| 3 | THE DESCRIPTION-EXPERIENCE GAP | 69 |
| 3.1 | Evidence for a behavioral gap | 69 |
| 3.1.1 | Two lines of research | 69 |
| 3.1.2 | First evidence | 70 |
| 3.1.3 | Testing the robustness of the description-experience gap | 74 |
| 3.2 | Objectives of the present work | 77 |
| 3.2.1 | First study | 77 |
| 3.2.2 | Second study | 78 |
| 4 | THE DESCRIPTION-EXPERIENCE GAP: A CHALLENGE FOR THE NEUROECONOMICS OF DECISION-MAKING UNDER UNCERTAINTY | 81 |
| 5 | THE IMPASSABLE GAP BETWEEN EXPERIENTIAL AND SYMBOLIC VALUES | 95 |
| 6 | DISCUSSION | 149 |
| 6.1 | Experiential and symbolic hybrid choices in previous literature | 152 |
| 6.1.1 | In monkeys | 152 |
| 6.1.2 | In humans | 155 |
| 6.2 | General considerations on the idea of a representational gap | 157 |
| 6.2.1 | Are value representations relative? | 158 |
| 6.2.2 | Ambiguity aversion | 160 |
| 6.2.3 | Fast-and-frugal heuristics | 164 |
| 6.2.4 | Policy-based models and value as a reification | 166 |
| 6.2.5 | Are values built <i>a posteriori</i> ? | 169 |
| 6.3 | Conclusion | 171 |
| | REFERENCES | 195 |
| | APPENDIX A APPENDIX | 197 |
| A.1 | Coordination over a unique medium of exchange under information scarcity . . . | 199 |
| A.2 | Interaction effects between consumer information and firms' decision rules in a duopoly: how cognitive features can impact market dynamics | 211 |

Listing of figures

| | | |
|-----|---|-----|
| 1.1 | The diamond-water paradox | 8 |
| 1.2 | Prototypical risk-attitudes | 16 |
| 1.3 | Prospect theory | 25 |
| 1.4 | Possible outcome sequences when a coin is tossed four times | 35 |
| 1.5 | Take the best heuristic | 37 |
| 2.1 | Classical and operant conditioning | 43 |
| 2.2 | Skinner’s box | 45 |
| 2.3 | The basic reinforcement learning components | 48 |
| 2.4 | Standard actor-critic architecture | 56 |
| 2.5 | Policy gradient learning | 60 |
| 2.6 | Raster plot showing how dopamine neurons encode prediction-error in classical conditioning | 64 |
| 2.7 | A valuation system for economic decisions | 66 |
| 3.1 | The continuum of uncertainty with regards to the description and experience paradigms. | 71 |
| 3.2 | Different types of outcome presentation in experience | 73 |
| 3.3 | A meta-analysis of the description-experience gap | 76 |
| 3.4 | RL model with descriptions of objective probabilities and payoffs provided to the individual | 79 |
| 6.1 | Experiential and symbolic hybrid choices in monkeys | 153 |
| 6.2 | Experiential and symbolic hybrid choices in humans | 156 |
| 6.3 | Ambiguity aversion pattern | 162 |
| 6.4 | Performance from the Experiential-Symbolic phase described in chapter 5, among experiments 1-to-8 | 165 |

Abbreviations

LTV: Labor Theory of Value

EUT: Expected Utility Theory

K& T: Kahneman and Tversky

VNM: Von Neumann and Morgenstern

BOLD: Blood Oxygen Level Dependent (signal)

CR: Conditioned response

US: Unconditioned response

CS: Conditioned stimulus

RL: Reinforcement Learning

fMRI: Functional Magnetic Resonance Imaging

RPE: Reward Prediction Error

OFC: Orbitofrontal Cortex

PFC: Prefrontal Cortex

ACC: Anterior Cingulate Cortex

ES: Experiential-Symbolic

EE: Experiential-Experiential

SP: Stated Probability

LE: Learning

E-option: Experiential Option

S-option: Symbolic Option

Active experimentation must force the apparent facts of nature into forms different to those in which they familiarly present themselves; and thus make them tell the truth about themselves, as torture may compel an unwilling witness to reveal what he has been concealing.

John Dewey, *Reconstruction in Philosophy*, 1920

La science est une construction qui fait émerger une découverte irréductible à la construction et aux conditions sociales qui l'ont rendue possible.

Pierre Bourdieu, *Science de la science et réflexivité*, 2001

O

Introduction

In the literature on value-based decision making, empirical measures of subjective values can be constructed from two sources: experience and description. Learning values from experience consists of acquiring information about the expected-value of an option via a trial-and-error process. In contrast, learning values from description requires to understand a symbolic language that will convey information about probabilities and outcomes. Both paradigms have been historically developed separately, and within different academic fields.

In order to understand how the construction of value is envisaged in each field, we will describe the historical events and the methodological specificities that allow its elicitation. It will hopefully

allow to shed light on our results, as well as contemporary debates regarding the ontological status (i.e. how is the construct of value materially translated in the brain) and epistemological role (i.e. to what extent this construct is useful) of subjective value for decision-making.

In the first chapter, we will discuss how subjective value is rooted in the notion of economic value, which most prominent classical economists first theorized as objective to further be interpreted in a subjectivist framework (utility). Furthermore, we will describe how we went from normative models of decision to descriptive models of decisions, by means of the description paradigm.

In a second chapter, we will see how psychology and neuroscience evolved toward integrating the notion of value to their model, notably through the paradigm of reinforcement learning. In addition, we will discuss how it led to the formulation of the classical two-step model of value-based decisions.

In a third chapter, we will introduce the *description-experience gap* phenomenon, that emerges from the meeting of the above lines of research.

The fourth and fifth chapters will include respectively, a literature review of the *description-experience gap*, and the main research paper of this thesis.

Lastly, the implications of the main experimental study will appear in discussion.

*But I have planted the tree of utility. I have planted it deep,
and spread it wide.*

Jeremy Bentham, *The Works of Jeremy Bentham*, 1843

*Nothing can have value without being an object of utility. If
it be useless, the labor contained in it is useless.*

Karl Marx, *Capital*, 1867

*How do human beings reason when the conditions for ratio-
nality postulated by the model of neoclassical economics are
not met?*

Herbert Simon, *The scientist as problem solver*, 1989

1

Utility theories and decision-making in economics

1.1 Toward a subjective theory of value

1.1.1 An objective theory of value: Labor theory of value

A fundamental tenet of classical political economics is that labor is one of the greatest determinant of economic value. This view was in fact already formulated by physiocrats¹: Richard Cantillon's

¹Physiocracy was a school of thought emerging in France in the late XVIII-th century, in the Age of Enlightenment. Physiocrats are often viewed as the founders of modern economic science, as well as economic liberalism, by promot-

Essai sur la Nature du Commerce en Général specifies that the ‘real or intrinsic value’ of a precious metal is ‘proportionable to the land and labour’ required for its production. However, characterizing economic variables and material processes — as opposed to ideal processes, value is thought to be a consequence of the material realm activity — which determine the objective (or normative) value of a good has given rise to series of dissensions among classical economic scholars. (King and McLure, 2014).

Adam Smith in *The Wealth of Nations* (1776), makes a clear distinction between the ‘real’ price (its ‘natural’ value) of a commodity and its ‘nominal’ price (Robertson and Taylor, 1957). Tackling the problem of inter-temporal variations in market prices, he states that labor is the ‘real standard’ and ‘real price’ by which commodities can be compared to one another and across time. In contrast, he states that ‘money is only their nominal price’ subject to volatility. Smith argues that labor acts a center of gravity offering a material ground for market prices, such that it is a reliable metric when it comes to market analysis. Among several arguments, Smith stressed that in a primitive state of society (where rents and lands are absent), there exists a necessary proportional relation between the exchange ratio of two goods and the quantity of labor necessary to produce them. In addition of labor, Smith also considered (in what we later called the cost-of-production theory of prices) various inputs as part of the output economic value, such as rents for instance.

David Ricardo (1835) completed Smith’s labor theory , notably by distinguishing the role of direct labor and indirect labor. Like Smith, he thought that prices could be explained by the quantity of labor incorporated in commodities. However, he notes that there exists a direct labor necessary to produce a commodity (e.g. workforce, tools), as well as an indirect labor, that is the labor producing the capital necessary for direct labor (e.g. the labor required to produce the tools). According to him, prices are governed by dynamics related to those two types of labors.

ing for instance the concept of “laissez-faire”. Their ideas influenced prominent classical economists, such as Adam Smith or David Ricardo.

While Smith and Ricardo described the underlying principles of the Labor Theory of Value (LTV), Karl Marx in *The Capital* (1873) went further to make it a cornerstone in analysis of capitalism. He thought of value as the origin of (conflicted) social relationships in the productive sphere. In Marx's thinking, value is derived from the labor time required of society for its formation. Formally: $W = C + L$, where W is the normative value (or worth) for a given product, C is the capital required in the process (e.g. machines, tools), and L is the quantity of labor. Moreover, value is fundamental to his sociological and political theory: social classes are determined in relation to the formation process of value. More precisely the position occupied by an individual in the productive sphere (either exchanging labor for a wage, or owning the means of production) defines the social class.

During the 19th century, the LTV was hegemonic, when it came to explaining economic value (Dillard, 1945). Importantly, Marx's writings and ideas became central to the European socialist movement, which assured the continuation of the LTV², at least in leftist political circles. However, in the academic sphere, the LTV was disputed and others favored a subjectivist conception of value: utility theory.

1.1.2 A subjective theory of value: utility

Although those three economists were proponents of an objectivist theory of value (or even an 'embodied' theory of value), they did not completely deny the role of utility (see Box. 1.1). Utility here refers to the subjective value or satisfaction one may experience from the consumption of a good. While they relativize its impact on prices, they nevertheless admit that utility is a necessary condition of economic value. As Ricardo notes, air and water have a greater utility than gold, while being less valuable. Then "utility is not the measure of exchangeable value [...] although it is absolutely essential to it". This paradox was also announced by Smith, and is known as the diamond-water paradox: "Nothing is more useful than water: but it will purchase scarcely any-

²Of note, even though the LTV will be disregarded by neoclassical economics, notorious economic theorists such as John Mayard Keynes supported the LTV (Keynes, 1936)

thing; scarcely anything can be had in exchange for it. A diamond, on the contrary, has scarcely any use-value; but a very great quantity of other goods may frequently be had in exchange for it.”. An answer to the diamond-water paradox has been given by W. Stanley Jevons (1871): While poorly valuable, water is of great utility. Yet once the first drink has been consumed, the marginal utility of water, which is very important when one is thirsty, decreases sharply so that the last drink has almost no value. Conversely, the marginal utility of diamonds (which involves for instance social prestige) decreases much more slowly (Fig.1.1).

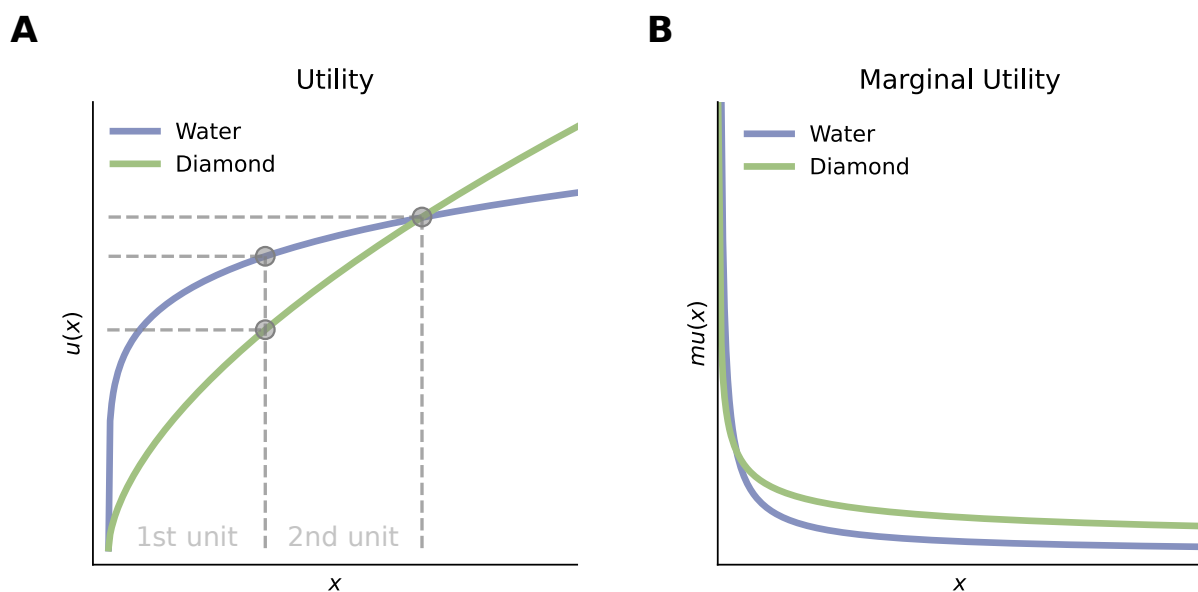


Figure 1.1: The diamond-water paradox. (A) Water has a higher utility than diamonds at modest levels of usage, making it more valuable, merely because people need it to survive. Yet water is available in large supply, and when it is consumed, its utility quickly decreases as its consumption is not urgent. In contrast, diamonds are in much lower supply and the urge for owning them is rather stable, such that the utility of one additional diamond becomes greater than the utility of one additional glass of water. (B) Hence the marginal utility (i.e., the utility gained when one unit is added) of water decreases much faster than the marginal utility of diamonds.

This solution illustrates the subjective conception of value Marginalism endorses. Marginalism posits that the value of a good is not determined by any inherent or intrinsic properties, nor by the amount of labor required for its production. Rather economic value is subjective, and is a proxy for the importance an individual puts into a good. Yet, while promoting it, the Marginalist movement is not the instigator of economic value understood as subjective utility. This conceptualization actually traces back to Bernoulli.

Box 1.1: Measurements in economics

One explanation for the reluctance toward utility theory (Viner, 1925) as a foundation of economic value might lie in measurement issues. In the 19th century, measuring incomes, prices, quantity of money... was common practice for economists. The labor theory of value was in the continuity of this tradition of measuring macroeconomic variables. In contrast, utility being a psychological phenomenon, quantifying it was an epistemological and technical challenge for economics at the time. Furthermore, at the end of the 19th century, the debate on the feasibility of psychophysics was raging (Moscati, 2018a). Gustave Fechner founded the discipline of psychophysics (Fechner, 1860), in an attempt to overcome the dualism between mental and physical substances. One way to realize that was to show that mental entities present measurable properties, possibly linked to physical phenomena. Along with pioneers of experimental psychology such as Wilhem Wundt, he claimed that sensations were measurable in a unit-based way, and were hence subject to scientific enquiry. French philosopher Henri Bergson (Bergson, 1889) disputed this claim, asserting that sensation are fundamentally qualitative. Consequently, he concludes that measuring sensations consists hence in a categorical error. Others, such that famous mathematician and philosopher of science Henri Poincaré (Poincaré, 1893), also opposed this view, asserting that unit-based measurement assumes transitivity, a property that sensations lack. Therefore, utility measurement, as the measurement of sensations in a unit-based way ^a was subject to identical criticisms, and faced identical epistemological challenges. Despite those attacks, utility theory became mainstream, notably with the foundation of neoclassical economics, in which the marginalist movement played a great role.

^aas utilitarian like Bentham defined it (Bentham, 1789)

1.1.3 The expected utility hypothesis

Bernoulli' proposed the first formalization of utility through expected utility theory (EUT) in his seminal 1738 paper (Bernoulli, 2011). Here he famously exposed a problem named the St. Petersburg paradox.

In this paradox, a casino offers to toss a coin over several trials. The initial stake begins at 2 ducats and is doubled each time the outcome is an head. Once a tail appears, the game ends and the player wins the accumulated monetary prize. In other words, the player wins 2 ducats if the first outcome is head, 4 ducats if head-tail, 8 dollars if head-head-tails, and so on.

What would be a fair price to enter the game? To answer this from the perspective of the gambler, we should consider the average payoff: winning 2 ducats has a probability of $\frac{1}{2}$, winning 4 ducats has a probability of $\frac{1}{4}$, and so on. According to a 1600s epistolary discussion between Blaise Pascal

and Pierre Fermat, a rational decision-maker should compute the mathematical expectation of each choice and then choose the option that yields the highest (Biswas, 1997). The expected value of this gamble X (i.e. the arithmetic mean of a large number of independent realizations of the gamble X) with n repetition is written:

$$E[X] = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + \dots + 2^n \cdot \frac{1}{n}$$

Said differently, the expected value converges to infinity, because the sum might grow endlessly. The paradox is that, according to the intuition³ of Bernoulli, there is probably of huge gap between what players would be willing to pay to play such a game (a few ducats), and its potential gains (infinity). To solve this paradox, Bernoulli introduced the utility function, as well as the presumption of a phenomenon called diminishing marginal utility. For Bernoulli, what matters to the player is the (expected) utility of the gamble, in other words the subjective and psychological anticipation of gains, not the expected values. One should thus first convert to subjective units, by means of a utility function:

$$u(X) = u(2) \cdot \frac{1}{2} + u(4) \cdot \frac{1}{4} + \dots + u(2^n) \cdot \frac{1}{n}$$

Furthermore, Bernoulli states that the shape of the utility function should be logarithmic, meaning it should marginally decrease (Fig. 1.1). Indeed, as he notes 'There is no doubt that a gain of one thousand ducats is more significant to the pauper than to a rich man though both gain are the same amount'.

1.1.4 Decision under risk

Bernoulli also set up the framework for decisions referred as 'under risk'. Risk means that outcomes are probabilistic events, as opposed to deterministic events. Later, Knight (1921) will distin-

³This intuition will be empirically verified later, revealing that most of the subjects are not willing to pay even 10 dollars to enter the game (Hayden and Platt, 2009)

guish risk and uncertainty. The former is quantifiable, under the form of a probability for instance. The latter expresses a situation where future events are essentially unpredictable due to the lack of any quantifiable knowledge. Another assumption made by Bernouilli, relies on a necessary maximization of expect value (or utility). This assumption is maintained within the framework of decision theory, even though many empirical and theoretical arguments will question its normative relevance (e.g. [Allais, 1953](#); [Sen, 1973](#); [Kahneman et al., 1991](#); [Tversky and Shafir, 1992](#)).

1.1.5 Moral utilitarianism and hedonistic utility

After Bernouilli, the construct of utility was echoed by the moral philosophy of utilitarianism, by British philosophers Jeremy Bentham (1789) and John Stuart Mills (1859). This theory assesses the morality of an action according to its consequences. Bentham defends that moral action maximizes a population pleasure (i.e. the sum of each individual utility) while Mills aimed for the minimization of pain. For Bentham, happiness as the maximization of aggregated utilities derives from his conception of the human being. An economic agent, a rational being, capable of calculation and logical reasoning for his personal case. Importantly, Bentham theorized the *felicific calculus*, an algorithm devised to compute the degree or amount of pleasure that a specific action is likely to induce. Several variables were included, such as the intensity (how strong the sensation is), duration (how long will it last), and the certainty (how likely will it occur). This sort of homoeconomicus⁴, that seeks to maximize its utility, is also evoked by Mill, although in 'altruistic' terms (see [Morgan, 2006](#)).

In the ethics of utilitarianism, utility thus refers explicitly to a form of hedonistic psychological phenomenon, which human beings are supposed to maximize (when positive), or minimize (when negative).

⁴Although the term is not used by Mill or Bentham, it is often admitted that they made one the first description of it ([Persky, 1995](#))

1.1.6 Marginal Revolution

On the basis of theoretical legacy of utilitarianism as well as Bernoulli's utility theory, a group of scholars from diverse countries led the paradigmatic shift, known today as the Marginal revolution. Instead of putting the emphasis on factors of production in the formation of value, they argued that economic value merely reflects and quantifies individual preferences (illustrated by marginal utility) given certain individual (i.e. psychological) and situational properties (e.g. being thirsty in an environment where water is lacking ensure that the latter will provide a great utility).

In great Britain, William Stanley Jevons posits that utility is the central calculus of economics. Via a series of articles culminating in his book *The Theory of Political Economy* (1871), he emphasized that "economic value depends entirely upon utility". In the legacy of the ethical utilitarian tradition (and especially Bentham's philosophy), he defended a quantitative and hedonistic vision of utility:

In the first place, pleasure and pain must be regarded as measured upon the same scale, and as having, therefore, the same dimensions, being quantities of the same kind, which can be added and subtracted. (Jevons, 1871)

He also predicted that although utility was not measurable at the time, it would soon, thanks to the quick development of scientific methods. He then proposed to consider utility as the 'willingness to pay', as an alternative.

In parallel, Carl Menger (1871) in Austria, and Léon Walras (1896) in France, independently developed different theories of utility. Menger, in *Principle of Economics* (1871), criticized Jevons for being too close to utilitarian hedonism (Jaffé, 1976), and thought that pleasure was to avoid as an economic object. Instead, he claimed that the need satisfied by the last unit of the good, that is the marginal utility of that unit, is what underlies subjective value. Contrary to Jevons, he did not take a stance on whether utility was or will be measurable in the future (Moscati, 2018b).

Léon Walras for his part, published *Elements of Pure Economics* (1896), where he distinguishes himself by an intense use of mathematics, which was uncommon in contemporary economics. Based on the concept of utility he demonstrated the existence of a general equilibrium in market

theory ⁵. He disagreed with Jevons on the idea that utility could be measured as the 'willingness to pay', as it depends on other factors such as the utility of other commodities and the individual's wealth. However, he agreed that utility should be indirectly measured, yet he did not provide any methods for doing so (Walras, 1909; Moscati, 2018b).

Jevons, Menger, and Walras were thus able to construct comprehensive theories of price and markets that quickly became popular among economists, making utility a foundational concept of economic science.

1.1.7 Ordinal Revolution

In Francis Ysidro Edgeworth 1881 work *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (1881), he presented a synthesis of utilitarianism (hedonistic utility) and psychophysics (Fechner, 1860), intending to make utility apt to observation. Economic theory consists in his view in a calculus of "hedonic forces", that is pleasure and pain. Economics investigation must then concentrate on unraveling the mechanism for pleasure maximization and conversely, pain minimization. According to Edgeworth, a pleasure could be measured incrementally. Said differently, it could be measured as perceivable increments in sensations, the scale starting from a zero-level (i.e. no stimulus). He labeled his discipline *hedonimetry*. In addition, the incremental approach proposed by hedonimetry suggested another property: ordinality. Indeed what mattered to Edgeworth, was the ranking of hedonistic sensations, not their absolute value. Therefore, he borrowed the distinction between ordinal and cardinal utility from economist Andreas Voigt (1893), who himself drew this distinction from Ernst Schröder ordinal and cardinal numbers (Schröder, 1873): Cardinal numbers, such as the number three, represent the total number of units that make up a given amount. Hence, cardinal numbers might be used for the absolute measure of a quantity. Ordinal numbers, on the other hand, come into play when counting the units that make up a quantity and represent the location or rank of a specific unit, such as the

⁵More precisely, he showed that supply and demand interact and tend toward a balance when an economy is composed by several markets working at once. Equilibrium theory will play an important role in modern macroeconomics (Arrow and Debreu, 1954)

third unit.

Translated to utility, ordinal utility represents the preferences of an individual if it is unique up to any (possibly non-linear), monotonic increasing transformation (Moscati, 2018b). Said differently, if $u(x)$ represents the individual's preferences, another utility function $u'(x) = F[u(x)]$, with F increasing, still represents the individual's preferences. In other words, it preserves the order between utilities.

In contrast, cardinal utility is more restrictive, as it requires only linear and positive transformation (Samuelson, 1938; Fishburn, 1970). Thus, only another utility function $\alpha u'(x) + \beta$, where $\alpha > 0$, maintains individual's preferences. Cardinal utility yet preserves more mathematical properties: the order between utilities as well as the order between utility differences. Psychologically speaking it suggests that cardinal utility allows to preserve the intensity of one's preference for one option over its alternative.

However, Edgeworth did not develop ordinal utility beyond hedonometry. In fact, The ordinalist revolution, which was mostly led by Irving Fisher (1907) and Vilfried Pareto (1897) originates from criticism relating to the psychological foundations of the principle of decreasing marginal utility, grounded in the framework of cardinal utility. Therefore, and Pareto in particular, pushed ordinal utility forward, in order to favor a more 'positive' approach to economics, freed from the psychological assumptions contained in cardinal utility ⁶.

1.2 Axiomatic utility

1.2.1 Revealed preferences

Through Samuelson *revealed preference theory* (1938), microeconomics furthered its 'escape from psychology' (Giocoli, 2005), by progressively eliminating psychological entities from theory. He aimed to break with the introspective psychological approach (which he considered non-empirical, as mental states and variables are assumed non-observable entities) inherited from moral utilitar-

⁶Of note, Pareto's positivism can be linked to Friedman subsequent views (Box 1.2; Serrano, 2006)

ianism and the early Marginalists. Samuelson thus argued that utility should be elicited directly from consumer choices (or even reduced to it, in order to get rid of the concept). An adequate empirical approach to the study of utility was then to infer it from preferences, which themselves are inferred from choices. By equating unobserved preferences with observed choices, *revealed preference theory* avoids circularity and make falsifiable predictions. Indeed, assuming people behave consistently and prefer option A to B, they should not prefer B to A thereafter. His approach to preferences measurements contrasted from endeavors to directly measure utility, for example through stated preferences or psychological methods. His aversion to psychological concepts also led him to be a proponent of ordinal utility (Moscatti, 2019).

1.2.2 Risk-attitudes

Mobilizing series of data collected from different institutions, Friedman and Savage (1948) made three observations that a robust theory of decision under risk (in their case EUT) should account for: (1) individuals of all income levels buy insurance ; (2) individuals of all income levels engage in gambling ; and (3) most individuals both purchase insurance and gamble. They ended up formalizing (in terms of deviation from a linear utility function) a set of different attitudes a decision-maker might adopt when faced to risky decisions. An individual is said risk-averse when its preference goes toward a safe option that provides systematically the same payoff instead of another option which outcome is a probabilistic event but has an identical expected value . The opposite behavior (preferring the uncertain option) is said risk-seeking. This risk typology is illustrated graphically by the curvature of the (cardinal) utility function, that can be concave (risk-aversion), convex (risk-seeking), or linear (risk-neutral) (Fig. 1.2). Risk-attitudes constitutes now a widely used framework, notably to characterize utility functions. Indeed, utility functions are mainly constructed by presenting risky gambles and analyze in which direction preferences tend (Vickrey, 1945).

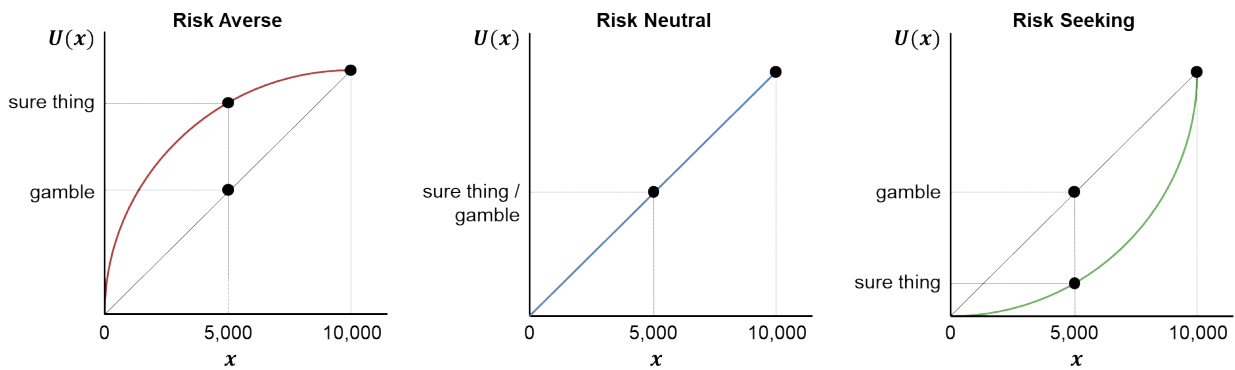


Figure 1.2: Prototypical risk-attitudes. The x -axis is in arbitrary units, such as a monetary reward. The y -axis represents the subjective utility function U for the x outcome. An individual that is said to be risk-averse (red curve) presents a concave utility function. Translated to behavior, its preference goes toward a safe option (with no outcome variance), as the safe option utility is overweighted, rather than a risky option (which outcome can vary). A risk-neutral individual (blue curve) presents a linear utility function, which means that the subjective valuation U does not affect the objective values of x . Hence, the individual is indifferent to risk, and will simply prefer the option with the highest expected-value. A risk-seeking individual (green curve) will present a convex utility function, and consequently favors risky gambles, as the safe option is underweighted. The figure is from [Bavard, 2021](#).

1.2.3 Von Neumann-Morgenstern utility theorem

From the 1930s to the early 1950s, expected utility theory came under severe criticisms ([Moscati, 2018b](#)). Notably, some economists suggested individuals might prioritize statistical properties of payoffs distributions (mean, variance etc.) rather than expected utility when making decisions ([Hicks, 1931](#)). Others noted that when decision is made under risk (meaning that payoffs and outcomes are governed by probability distributions), payoffs as well as the utility derived become random variables, which implies that utility is cardinal, contrasting with the contemporary ordinal conception of utility ([Tintner, 1942](#)). Others were proponents of simpler decision-rules such as “minimax” ([Wald, 1951](#)), or the mere idea that decision-makers focus on extreme outcomes to assess risky choices ([Shackle, 1949](#)).

Despite those attacks, cardinal EUT became the most important foundation for decision under risk, ruling out⁷ the ordinal approach, with the robust axiomatic provided by John von Neumann and Oskar Morgenstern (VNM).

In their book *Theory of Games and Economic Behavior* (1944), they introduce a set of axioms under-

⁷Cardinal utility seems to be dominant in contemporary decision under risk and other areas of microeconomics. However, ordinal utility is still used in demand analysis for instance ([Moscati et al., 2013](#)).

lying agents' decisions, and ensuring their rational behavior. A novelty of their approach consists in the analysis of decision within a game framework, where rationality implies the anticipation of others behavior. This abstract model of the rational agent is often called *normative expected utility theory* (as opposed to descriptive). The ensuing decision axioms prescribe how an hypothetical agent *should* behave given consistent rationality axioms, in order to further derive a utility function. Thus, VNM assumed an agent with fixed, well-ordered preferences, that has 'perfect information', and behaves 'as-if' (Box. 1.2) it maximizes expected-value.

Their first axiom requires that the preferences are complete. Let's imagine two lotteries L_1 and L_2 : either L_1 is preferred to L_2 , either L_2 is preferred to L_1 . The case where both statements are true is allowed, it merely means that the decision-maker is indifferent between the lotteries.

Axiom 1: Completeness

For lotteries L_1, L_2 , either $L_1 \succ L_2$, either $L_2 \succ L_1$, or $L_1 \sim L_2$

The second axiom is also rather basic. It states that preferences are transitive. In other words, if L_1 is preferred to L_2 and L_2 is preferred to L_3 , then consistent preferences suppose L_1 is preferred to L_3 .

Axiom 2: Transitivity

For lotteries L_1, L_2, L_3 , if $L_1 \succ L_2$ and $L_2 \succ L_3$ then $L_1 \succ L_3$

The next axioms are more technical in nature as they suppose compound lotteries. A compound lottery consists of running a random device which yields other lotteries and not a monetary prize. Let's suppose two lotteries L_1 and L_2 . We then run a compound lottery which may result in two outcomes: A , with probability p and B with probability $1-p$. When A is drawn, the decision-maker obtains the outcome of L_1 ; conversely if B is drawn, the outcome of L_2 is obtained. It follows that if L_1 is strictly preferred to L_2 , then there exist a value for p where $pL_1 + (1-p)L_2$ is also preferred

to L_2 . Said differently, as p gets closer to 1, the compound lottery $pL_1 + (1 - p)L_2$ gets similar to L_1 , which at some point leads to a strict preference of $pL_1 + (1 - p)L_2$ over L_2 .

Axiom 3: Continuity

For lotteries L_1, L_2 if $L_1 \succ L_2$ then for some number $p \in [0, 1]$

$$pL_1 + (1 - p)L_2 \succ L_2$$

Lastly, the independence axiom (also named independence to irrelevant alternatives) assumes that a third L_3 has no impact on the above preference relation. Using the same example as above while adding two compounds: 1) if A is drawn, we obtain the outcome of L_1 , if B is drawn we obtain the outcome of L_3 2) 1) if A is drawn, we obtain the outcome of L_2 , if B is drawn we obtain the outcome of L_3 . It follows that $pL_1 + (1 - p)L_3$ is preferred to $pL_2 + (1 - p)L_3$. In both B situations, we obtain L_3 , then remains the A situations. Assuming a strict preference of L_1 over L_2 we should prefer the first compound to the second.

Axiom 4: Independence

For lotteries L_1, L_2 , and any lottery L_3 , if $L_1 \succ L_2$ then for any number $p \in [0, 1]$

$$pL_1 + (1 - p)L_3 \succ pL_2 + (1 - p)L_3$$

Finally, a utility function u is said to possess the expected utility property if, for a gamble X , which yields n outcomes x_i with n associated probabilities p_i :

$$u(X) = p_1u(x_1) + p_2u(x_2) + \dots + p_nu(x_n)$$

The strength of the VNM utility theory lies in the demonstration of the existence of such a utility function, as long as the preference axioms are all satisfied. However, several weaknesses

rapidly arise regarding the behavioral validity of the VNM utility theorem, consequently questioning whether or not it should be used as a normative model for decision theory (Box 1.3).

Box 1.2: As-if hypothesis and instrumentalism

Friedman and Savage (1948) proposed an anti-realist vision of EUT (Wong, 1973; Boland, 1979). The counterpart of anti-realism is scientific realism, a philosophical position in the epistemology of science that can be broken down into three claims (Chakravartty, 2017):

- Metaphysical claim: The external world is ontologically separated from the mind, it exists independently of one's lived experience.
- Semantic claim: Whether true or false, statements about scientific entities (observable and unobservable) are truth-apt.
- Epistemological claim: Theoretical statements (describing a mind-independent reality) constitutes knowledge of the world. Truth consists in a relation to reality, a statement that is meaningful and truth-apt must correspond to an entity in the external world (i.e. correspondence theory of truth).

Even though Savage and Friedman grant the metaphysical commitment, they sort of oppose the two following claims, by promoting an instrumental epistemology, where descriptions of unobservables merely are instruments for the prediction of observable phenomena, implying that those descriptions are not intended to be true. Applied to EUT, it states that individuals consciously calculating expected utilities is not what should concern economists. Rather, individuals should be considered as behaving "as-if" they calculated and compared (unobservable) expected utilities. This instrumental epistemology will culminate in Friedman's methodological essay (Friedman, 1953) where he asserts that the accuracy of the model predictions should prevail on the realism of the theoretical assumptions. In his epistemology, Friedman is thus agnostic regarding the implementation of utility and its computation: It is simply a scientific construct that should be used as a tool to make predictions about economic behavior. This attitude consisting in the absence of ontological commitments with regards to utility will typically be adopted by most economists (Gul and Pesendorfer, 2008).

1.2.4 Allais paradox

A few years after the publication of *Theory of Games and Economic Behavior*, in 1952, a conference focusing on decision under risk was held in Paris. Maurice Allais with a group of French economists openly challenged the proponents of the expected utility hypothesis. Allegedly, during a conference break, Allais exposed a gambling problem to Leonard Savage⁸ (Moscati, 2018b), which

⁸Savage was a statistician particularly interested in decision theory, as well as a fierce defender of EUT at the time. In later works such as *The Foundations of Statistics* published in 1954, he proposes a subjective theory of probability (or subjective expected utility theory) which became a cornerstone for Bayesian inference in game and decision theory.

will be later known as the Allais paradox⁹. The problem starts with a first gamble:

Allais paradox: Gamble 1

- (L_1) a safe option yielding 1 million for sure.
- (L_2) a risky option that yields 5 millions with probability 0.1, 1 million with probability 0.89, and 0 with probability 0.01

Allais argued that L_1 was highly appealing and more prudent, to which Savage agreed. He then proposed a second gamble:

Allais paradox: Gamble 2

- (L_3) a risky option yielding 1 million with probability 0.89, and 0 with probability 0.11.
- (L_4) a risky option yielding 5 millions with probability 0.1, and 0 with probability 0.9

Strikingly, Savage here preferred L_4 to L_3 . Allais remarked that Savage had just violated EUT. Indeed, to be normatively valid, the pair of preference should follow a unique utility function. If $L_1 \succ L_2$ it implies that $u(1) > 0.1 \cdot u(5) + 0.89 \cdot u(1) + 0.01 \cdot u(0)$. Yet, if $L_4 \succ L_3$ it means that $0.1 \cdot u(5) + 0.9 \cdot u(0) > 0.11 \cdot u(1) + 0.89u(0)$. But there exists no utility function satisfying both inequalities: either $L_1 \succ L_2$, implying that $L_3 \succ L_4$, or $L_2 \succ L_1$ and therefore $L_4 \succ L_3$.

In other words, the inconsistency stems from the violation of the independence axiom. Another way to understand this paradox is to rewrite L_1 and L_4 . L_1 and L_2 can be both seen as offering an outcome of 1 million with probability 0.89. Also, both L_3 and L_4 give an outcome of nothing with probability 0.89.

Allais paradox: Gamble 1' and 2'

- (L_1) yields 1 million with probability 0.89, and 1 million with probability 0.11.
- (L_2) yields 5 millions with probability 0.1, 1 million with probability 0.89, and 0 with probability 0.01
- (L_3) yields 1 million with probability 0.89, and 0 with probability 0.11.
- (L_4) yields 5 millions with probability 0.1, 0 with probability 0.89, and 0 with probability 0.01

⁹The paradox appears in the essay published later to the conference, in French and in *Econometrica*(Allais, 1953)

Now when disregarding the probability 0.89 (i.e. considering it as an irrelevant alternative) and consequently equalizing the outcomes, L_2 has a probability of 0.01 to win nothing and a probability of 0.1 to win 5 millions. L_4 has identical contingencies. In the same manner, L_1 and L_3 become the same choice. Therefore, the choice pair $(L_1 \succ L_2, L_4 \succ L_3)$ violates EUT independence and is normatively irrational.

Box 1.3: Allais normative and experimental model of utility

After the Paris episode, Allais distributed a questionnaire by post to the participants of a seminar he was conducting. The questionnaire undoubtedly included Allais paradox and perhaps counterexamples, albeit it was not its objective to test the paradox. Its main purpose was to empirically characterize the utility functions of the participants (Mongin, 2019). He considered the VNM axiomatic, as not satisfying the properties of measurability that one expects of a utility function. As a proponent of early cardinalist theories, he thought that no preferences could be properly derived from a utility function that was not able to measure the intensity of a preference over another. Indeed, unlike other VNM theorists, he denied that a utility function satisfying VNM axioms would necessarily provide such measurement. His empiricist stance might also be opposed to Friedman's epistemology, where economic models do not aim at describing plausible decision processes. Thus, Allais paradox is remembered as an empirical refutation of EUT and especially of the VNM axiomatic. However Allais was actually aiming at proposing a normative countermodel of the 'rational man', which would be based on the rationality empirically observed in human subjects.

1.3 Behavioral models of value and decision-making

1.3.1 Anomalies in decision-making

The above outlined attack heralded a series of empirical findings showing that under certain circumstances, human subjects tend to deviate from EUT predictions, and make choices violating VNM axiomatic (e.g. Ellsberg, 1961; Kahneman and Tversky, 1972; Bell et al., 1988; Tversky and Shafir, 1992).

An important result from this period is that people's preferences are constructed in the process of elicitation, and consequently elicited values and preferences are highly dependent on measurement methods. Indeed, normatively equivalent methods of elicitation often produce system-

atically different responses (Slovic and Lichtenstein, 1971; Slovic, 1995; Lichtenstein and Slovic, 2006). These *preference reversals* may occur when subjects are confronted with two lotteries: L_1 which offers a large sum of money, but associated with a relatively small probability of winning, and L_2 which offers less money, but with a greater probability of winning. Subjects are then asked to perform two tasks: a choice between L_1 and L_2 , and thereafter attach a certainty equivalent to each prospect, i.e. a fictional alternative lottery certain enough such that they would definitely choose it. A typical finding is that subjects prefer L_1 when choice elicited, while paradoxically, L_2 is given the higher valuation when it comes to certainty equivalents. These *preference reversals* thus explicitly violate the principle of procedure invariance that is fundamental to theories of rational choice and raises difficulties regarding the elicitation of preferences and thus utility. A common interpretation is that preference reversals are evidence for the existence of several systems of preferences (Slovic and Lichtenstein, 1983; Tversky et al., 1988).

1.3.2 Judgment and Decision-Making

One task for decision theorists was then to amend normative models (such as EUT) and propose empirical models of utility (or descriptive models), while looking for systematic deviations. Those systematic deviations from optimality such as defined by normative models (which themselves are constructed by the use of mathematical or philosophical arguments) are called biases. When biases are found, it can fuel the creation of novel descriptive models, that better account for the observed departures from the norm, often with the language of cognitive psychology (Baron et al., 2004). Based on the conjunction of normative and descriptive models, prescriptive models can be proposed, possibly to improve applied decisions (Kahneman et al., 1982a; Leonard, 2008). This basically outlines the three-way model on which the nascent field of judgments and decision-making (JDM) is based (Freeling, 1984; Baron, 1995; Bell et al., 1988).

1.3.3 Different research programs

Mostly emerging in the 1970s, a structuring element of this literature was the separation between topics related to psychology (biases and heuristics) and topics related to economics (risk, uncertainty and utility models). Those two lines of research led to two kinds of descriptive models of decision making. Namely, heuristics and utility models. Heuristics tend to be less formalized and complex than utility models, by supposing simple rule-based psychological operations instead of utility computation. One could then see those utility-free models as more parsimonious (Epstein, 1984), as they also relax rationality assumptions. There are nonetheless two approaches to heuristics, roughly identified with the one associated with the Kahneman and Tversky's heuristics-and-biases program and Gigerenzer's fast-and-frugal-heuristics program. The scientific dispute here lies within the appropriate normative standard for judging human decision-making (Gigerenzer, 1996; Vranas, 2000; Polonioli, 2013).

In contrast, the risk and uncertainty program involves what we will call here the *description paradigm of decision-making* (Box 1.4), where although arrangements are made with the assumption of fixed, well-ordered preferences (models are more flexible) and 'perfect information' (experiments are designed to provide as much a priori information as possible), the notion of value computation as well as a high level of formalism are still maintained. They are, in this regard, the direct continuation of previously seen studies of utility.

Thus, we can distinguish three modeling approach to decisions under uncertainty: (i) models with value (or utility) computation for isolated options (ii) models with value-difference computation, meaning that an option is assessed relatively to another (iii) utility-free heuristics, which explain decisions in terms of computational shortcuts.

In the below section, I sought to present one illustrative example for each approach, although there exist numerous other models (Vlaev et al., 2011).

1.3.4 Value models

Value-first models: the case of prospect theory

Prospect theory (PT) was developed by Daniel Kahneman and Amos Tversky (K&T) in 1979 (Kahneman and Tversky, 1979), and earned Daniel Kahneman the Nobel Prize in Economics in 2002. This theory is considered foundational of behavioural economics and is one of the first occasions utility theory was based on experimental work. In their experiment, K&T presented several prospects to their subjects. These gambles typically take the form of a choice between two lotteries (presented textually), for instance:

Prospect theory: Gamble 1

- (L_1) a risky option with probability .33 of winning 2500, probability .66 of winning 2400, and 0 otherwise.
- (L_2) a safe option that provides 2400 with certainty.

Varying gambles including risky and safe options in the gains and losses domains, they observe that subjects present an asymmetric value function. Behaviorally translated, subjects will typically be risk-averse in gains (preferring the safe option L_1) while they often display risk-seeking preferences in losses (thus choosing the risky option L_2 when outcomes are framed as negative). This behavior results in an inferred S-shaped utility function (Fig. 1.3).

This particular shape illustrates the *reflection effect*, which consists of opposite risk-attitudes articulated around a reference point. Indeed, gains and losses domains are not here to be understood in absolute terms, but rather relatively to this reference point, which is thought to be set by the subject subsequently to a few gambles presentations. In addition, PT predicts another phenomena: loss aversion. Loss aversion furthers the idea that losses are treated differently to gains. It is implemented by including a factor that amplifies negative outcomes. Formally, the utility function curvature for an outcome x is thus defined as follows:

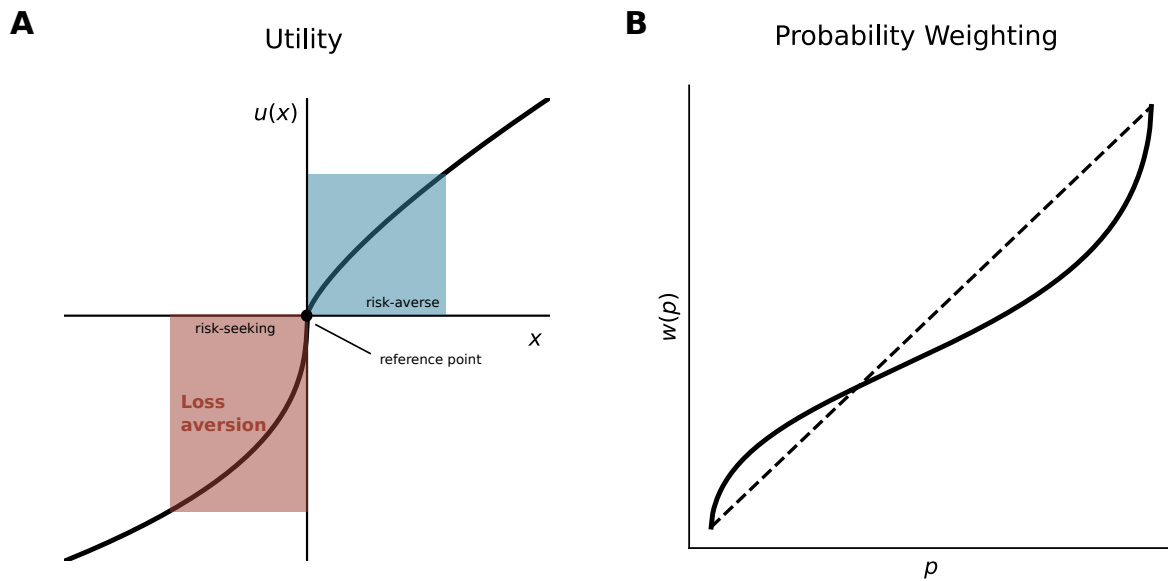


Figure 1.3: Prospect theory. **(A)** A typical shape for the PT value function u is an asymmetrical s-shaped, also known as the *reflection effect*. This shape is articulated around a reference point, which means that gains and losses are to be understood relatively to the range of possible outcomes. The curve is steeper and convex in losses (risk-seeking), a phenomenon called *loss aversion*, which translates psychologically as an overweight of negative outcomes. Conversely in gains, people are generally risk-averse, displaying a concave curve. **(B)** Empirical characterization of the probability weighting function w most of the time results in an inverse s-shape. Hence, low probabilities are overweighted, while middle and high probabilities are underweighted.

$$u(x) = \begin{cases} x^\alpha & \text{if } x > 0 \\ -\lambda(-x)^\beta & \text{if } x < 0 \end{cases}$$

with λ being the loss aversion parameter, that increases the steepness of the loss curve. A value of $\lambda > 1$ supposes loss aversion (typical empirical value oscillates around 2). A value of $\beta < 1$ means risk-seeking attitudes while $\beta > 1$ corresponds to risk-averse attitudes. In the gain domain, this relationship is identical regarding the values of α . A decision-maker with an $\alpha < 1$, a $\beta < 1$ and $\lambda > 1$ is prototypical of what K&T observed in their study: presenting risk-averse attitudes in gains, while being risk-seeking in losses in addition of being loss averse.

Another core feature of PT is the probability weighting function (Fig. 1.3B):

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}$$

with γ controlling the curvature. When $\gamma = 1$ the function is linear (meaning there are no probability distortions). When $\gamma > 1$, the function tends progressively toward an s-shape. Conversely, when $\gamma < 1$ the curve adopts an inverse s-shape. According to K&T, this component of PT is crucial as the subjective distortion of probabilities provides an explanation for Allais' paradox. Indeed, objective probabilities are weighted according to 'the impact of events on the desirability of prospects and not merely the perceived likelihood of these events' (Tversky and Kahneman, 1979). A classical result is the overweighting of small probabilities while high and intermediate probabilities are underweighted. This property can cause the sum of weighted probabilities to equal less than one. According to K&T this 'subcertainty effect', explains the violation of EUT in Allais' case.

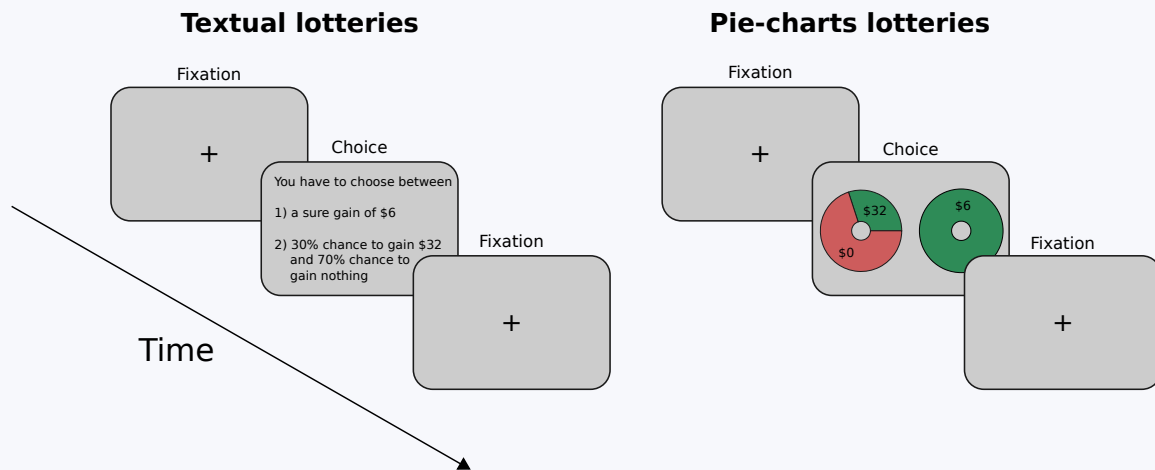
Finally, the subjective expected utility U of a gamble X , is computed as the sum of the utility of n possible outcomes weighted by their associated subjective probability:

$$U(X) = \sum_{i=1}^n u(x_i) \cdot w(p_i)$$

Prospect theory is still among the most influential theories of decision under risk nowadays. It has been cited as an explanatory framework for a broad range of behaviors (Barberis, 2013): in finance (Baker and Nofsinger, 2010), housing investment (Genesove and Mayer, 2001) or even political conflicts (Levy, 1996). Following its publication, several criticisms arose, notably that prospect theory is agnostic regarding the core decision processes, and therefore does not inform us about cognition *per se* nor provides mechanistic explanation for the psychological phenomenon it describes (Trepel et al., 2005; Barberis, 2013). Also, some have suggested that simpler decision rule could make a better account (in a parsimonious manner) of the predictions made by prospect theories (e.g. Brandstätter et al., 2006). Yet, prospect theory remains among the most cited papers in the JDM literature, while being epistemically powerful (from a falsificationist perspective) as it is highly replicable (Ruggeri et al., 2020).

Box 1.4: The description paradigm of decision-making

The gambling metaphor of individual choice (Goldstein and Hogarth, 1997) suggests that gambles work as a prototypical abstraction for real-life decisions (Savage, 1954): an act (a choice between alternatives) leads to multiple consequences (outcomes), which themselves are probabilistic event (probabilities) s. The experimental paradigm of decision under risk thus consists in choices between ‘lotteries’ or ‘gambles’, i.e., options associated to (most of the time) known probabilities and outcomes.



Gambles in experimental settings are often represented as lotteries described textually (e.g. (Tversky and Kahneman, 1992b)) or as pie-charts (e.g. De Martino et al., 2006). Also, in most studies, decision problem are single-shot (i.e. a gamble is only showed once), and the outcome of the choice is often not presented. Choices among these lotteries reveal risk preferences, loss aversion, and ambiguity aversion (Samuelson, 1938; Holt and Laury, 2002; Varian, 2006) and the data serve as input for decision or utility models and parameter estimation procedures. The best fitting parameter(s) are identified via a statistical estimation procedure, often maximum likelihood estimation (Harless and Camerer, 1994; Regenwetter and Robinson, 2017; Harless and Camerer, 1994). A perfectly parameterized model is supposed to reproduce 100% of an individual choice history. Going through an individual choice history, we can estimate $P(\Theta|D)$, i.e. the likelihood of a specific choice under certain parameters' values, with Θ being the set of free-parameters (e.g. the loss aversion parameter from Prospect Theory) and D the data (the choice). Thus, the statistical optimization consists in maximizing the sum of $P(\Theta|D)$ for all decisions within an individual choice history.

Comparison-based with value-computation models: Regret theory

Regret theory (Loomes and Sugden, 1982; Loomes et al., 1992) was based on the intuition that rather than evaluating prospects in terms of a summary statistic for each option individually (as EUT, or Prospect theory), decision-makers are concerned with state-contingent payoffs, i.e. the difference between options' payoffs within a certain state. An implication of this model is that individuals

rather than maximizing absolute expected-values, seek to minimize the regret resulting from a low payoff choice, when a better alternative yielded higher payoff. Conversely, an individual should maximize the rejoicing that arises when a choice is indeed optimal.

Formally, regret theory assumes both a finite state space of $S = \{s_1, \dots, s_n\}$, and probabilities $P = \{p_1, \dots, p_n\}$. An action is a function A that maps $A(s) \rightarrow X$ where X is an outcome set (e.g. money amounts). If we suppose 2 actions, denoted by $x_{1,s}$, the outcome of action A_1 in state s_i is realized with the probability p_i in that state. If state s is realized, and considering a choice of A_1 over its alternative A_2 , the decision-maker receives outcome $x_{1,s}$, while the alternative choice would have yielded $x_{2,s}$.

Considering the utility function u , the preference relation between the two actions A_1 and A_2 , for all states s_i is thus expressed as follow:

$$A_1 \succsim A_2 \Leftrightarrow \sum_{i=1}^n p_i \cdot u[A_1(s_i)] \geq \sum_{i=1}^n p_i \cdot u[A_2(s_i)]$$

Thus, according to regret theory, a decision-maker would seek to minimize the regret (or maximize the rejoice) by maximizing the difference of the yielded utilities across all states, i.e. by discriminating the best from the worst action in terms of relative payoff. An interesting feature of regret theory is that it provides a rational for preference reversals as well as Allais paradox (Bleichrodt and Wakker, 2015). In short, if the information processing is different in choice elicitation (value difference) compared to directly stated valuation (absolute value) it can explain why different elicitation methods yield different preferences.

Utility in the brain?

Camerer, Loewenstein, and Prelec (2004a) proposed the neuroeconomics agenda: to apply neuroscience techniques and expertise to economic research. At the time, researchers hoped that the new functional magnetic resonance imaging (fMRI) technology would allow them to pinpoint which parts of the human brain are engaged in various sorts of economic decisions. Furthermore, fMRI techniques were also seen as useful tools to adjudicate between different descriptive mod-

els. For instance, the well known PT "loss aversion" was characterized neurally, by first showing that regions such as the ventral striatum presented "neural loss aversion". Indeed the decrease in activity for losses was steeper in that region, than the increase in activity for gains (Tom et al., 2007b). Other components of PT, such as probability weighting (Paulus and Frank, 2006) or the framing effect (De Martino et al., 2006), were also found to potentially have neural representations. Similarly, comparison-based utility models (such as regret theory) found empirical support, when neurophysiological recordings in monkeys and humans shown comparative reward coding in neural substrates (Nieuwenhuis et al., 2005; Tobler et al., 2005). The neural determinants of subjective valuation are further discussed in chapter 2.

1.3.5 Comparison-based and value-free models

Bounded rationality

In the 1970s, bounded rationality (Simon, 1955) emerged as an alternative basis for the formalism proposed by EUT and neoclassical economic modelling of decision-making. Simon goals was to propose a theory of the rational choice that was compatible with the limited nature of human's cognition. In short, contrary to the neoclassical model that assumed a decision-maker fully informed with infinite cognitive abilities, bounded rationality assumes an agent with limited computational resources which makes cost-efficiency trade-offs in a complex environment where information is lacking. In addition, he fiercely criticized the assumption of expected value (or utility) maximization. According to him, a boundedly rational agent attempts to attain some satisfactory or sufficient outcome, but not necessarily an optimal or maximal one (Simon, 1947, 1972).

Simon also highlighted the contribution of learning, perception, and other cognitive processes in decision-making.

In Savage's world framework, small worlds are to be distinguished from large worlds (Savage, 1954). An environment with perfect and full information is called a small world, while a large word supposes that the relevant information is unknown or has to be estimated from few obser-

vations, so that the requirements for rational decisions are a priori not met. Savage and Simon both emphasize that in large worlds one can no longer expect standard models of rationality to provide the correct answer.

Heuristics-and-biases program

The term “heuristic” originates from mathematician George Polya (1945), who was trying to describe to its students how mathematicians reason. A heuristic is close to an algorithm, constituted by a set of rules, executed in sequential order. Algorithms however are to achieve a certain goal, with clear conditions. By contrast, a heuristic is a rule without very clear conditions, and does not necessarily results in a useful outcome. Its role is to suggest another approach to a problem, which may subsequently lead to a solution. For instance, when facing a mathematical problem: Are there analogous problems? Can you use schematics to represent the problem? Etc.

Kahneman and Tversky (1972) took up the idea of heuristics to explain biases in probability judgment. Taking Bayesian probability theory as a normative model, they found that people made judgments that were inconsistent with Bayes’ theorem. Indeed, following Savage’s SEU (Savage, 1954), an extensive literature tested the idea of the individual as a Bayes’ decision-maker (Slovic and Lichtenstein, 1971). These findings have fostered an approach where the decision-maker is viewed as a conservative Bayesian estimator, within which departures from the norm are attributed for instance to a deformation of some evidence (extreme ones in a particular). This conservatism implies that subjective posterior estimates should be monotonically related to objective Bayesian values (Edwards, 1968). To test the robustness of those previous results, K&T presented sampling problems of the following form:

Representativeness heuristic: Problem 1

Consider two very large decks of cards, denoted A and B.
 In deck A, $\frac{5}{6}$ of the cards are marked X and $\frac{1}{6}$ are marked O. In deck B, $\frac{1}{6}$ of the cards are marked X and $\frac{5}{6}$ are marked O. One of the decks has been selected by chance, and 12 cards have been drawn at random from it, of which 8 are marked X and 4 are marked O.
 What do you think the probability is that the 12 cards were drawn deck A, that is, from the deck in which most of the cards are marked X?

The above problem was alternatively presented with proportion $\frac{5}{6}$ replaced by $\frac{2}{3}$, and $\frac{1}{6}$ by $\frac{1}{3}$. The ratio of drawn cards (8:4 here) was also varied. K&T observed that when they asked subjects for the odds that the cards were drawn from one of the two proposed decks, subjects ignored proportion differences ($\frac{5}{6}, \frac{1}{6}$), and relied predominantly on drawn cards ratio. Obviously, from a normative standpoint, the proportion of each kind of marked cards initially present in the deck have a substantial impact on objective posterior odds. Subjective posterior estimates however diverge greatly, to the extent that they are not even monotonically related to objective probabilities. To explain those results, K&T assumed that people were not attempting to apply Bayes rule. Rather, they propose that they apply a heuristic of similarity, where the ratio of drawn cards (sample) is compared to the ratio of initial decks (population), without however taking into account the actual size of the initial decks. While being relevant, this similarity, or representativeness heuristic, is qualitatively different from Bayes computation of posterior probabilities. Bayes' theorem states that:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

where H stands for the hypothesis (those cards come from deck A) and D the data (sample of drawn cards). In the case of the representativeness heuristic, it becomes:

$$P(H|D) = P(D|H)$$

which constitutes a violation of the theorem.

A more illustrative example of the representativeness heuristic consists of a task where K&T pre-

sented personality sketch of a graduate student (named Tom) to subjects, which was conceived to match the stereotype of a computer science student (Kahneman and Tversky, 1973). Thereafter they asked participants to rank various academic fields according to the likelihood that the fictive student belongs to one of them. Earlier, K&T had asked other subjects to what extent Tom is representative of the prototypical graduate student of several study areas. K&T observed that these rankings were strongly correlated with each other. In other words, subjects inferred that Tom was a computer science student because his traits matched the stereotype. However, subjects were informed that computer science was a small field at the time, which counted few students. Although this information reduces the likelihood that Tom is a student in that field, they ignored that fact when making their judgments. K&T concluded from this evidence that people were using a representativeness heuristic, leading to a bias with regards to the normative model, which would take into account both population size and similarity.

Following this study and using similar experimental methods, K&T identified a series of heuristics, such as:

- **Availability heuristic** (Tversky and Kahneman, 1973): People assess the frequency of events by availability, i.e. by the ease at which it comes to mind. It leads people to overestimate the likelihood of an event solely because it can be recalled quickly.
- **Anchoring and adjustment heuristic** (Tversky and Kahneman, 1974): When asked for estimating unknown quantities, people tend to start with information one does know (the anchor) and then adjust until an acceptable value is reached. This heuristic can lead to biases when the anchor is not relevant to the considered problem. However, the anchoring and adjustment process can also help producing estimates closer to optimality when the initial information is relevant (e.g. When was George Washington elected? You can quickly generate an estimate by adjusting from the date of the Declaration of Independence in 1776, a date known to be close to the correct answer).
- **Simulation heuristic** (Kahneman and Tversky, 1981): Assessments of propensity or like-

likelihood of an event are derived from mental simulations. Among other biases, people seemingly experience more regret over outcomes that are easier to imagine, easier to 'picture in mind'. According to K&T, people for instance use this to answer questions involving counterfactual or causal propositions.

For K&T, heuristics were then useful as they saved time and cognitive resources, but also because they could sometimes provide quasi-optimal answers. Nonetheless, heuristics were first invoked to explain biases, i.e. departures from normative models. Even if heuristics were functional, they still led to suboptimal choices and errors in important and possibly frequent situations, such as in medical diagnosis (Kahneman et al., 1982b). Subsequently, Kahneman & Frederick (2002) will propose a definition for heuristics:

An heuristic assesses a target attribute by another property (attribute substitution) that comes more readily to mind.

Fast and Frugal heuristics program

Despite being highly influential; one could argue that the influence of the heuristics-and-biases program declined in the 2000s, at least in psychology (McKenzie, 2005; Truc, 2021). A possible origin of this decline can be found in the critique of the approach by the fast-and-frugal research program (Gigerenzer, 1991, 1996). This critique can be summarized in three main arguments:

- explaining cognitive phenomena using labels such as availability and representativeness is vague, and says nothing about the processes underlying judgment. Additionally they prevent the development of comprehensive theories of decision-making (Gigerenzer, 1996).
- focusing mainly on coherence standards (i.e. quantifying the deviation from a normative statistical model) leads to the pitfall of neglecting the role played by the environment, and how individuals rationality can be assessed relatively to the environment ('ecological rationality') (Gigerenzer and Gaissmaier, 2011)

- consequently, the approach favors the building of 'biases lists' and leads to the 'bias-bias', i.e. seeing systematic errors where in fact ecological rationality is applied. This has undesirable effects on public policies inspired by the heuristics-and-biases literature (Gigerenzer, 2018).

The “vagueness” argument has been illustrated using two related phenomena: the gambler’s fallacy and the hot-hand (Gigerenzer and Brighton, 2009). Both phenomena are rooted in intuitions about randomness. We typically observe the gambler’s fallacy when people intuitively predict that in a binary outcome sequence, after a long run of one outcome, the other outcome must appear. A classic example of this fallacy happens when we flip a fair coin: people have a tendency to predict heads after a sequence of tails (Kahneman and Tversky, 1972). In contrast, the hot-hand fallacy refers to the tendency that a sequence of identical outcomes will continue. A classic example consists in the prediction that a basketball player will score again after a succession of baskets (Gilovich et al., 1985). In the case of the coin, short sequences are believed to be “representative” of their generating (random) process, whereas a player scoring several points in a row is perceived as good (and thus reinforce the hypothesis that his performance is not due to chance) leads to predicting a continuity. Those two phenomena have thus, been designated as the consequence of the representativeness heuristic (Ayton and Fischer, 2004). However, the 'representativeness' label can be seen as vague, as opposite outcomes are explained without specifying the underlying mechanisms leading to such a prediction. If the priors leading to such predictions are not included in the heuristic formulation, then it is incomplete. One could thus see it as epistemologically 'weak' in terms of both explanatory and predictive power (Gigerenzer, 1996).

Moreover, it can be argued that these intuitions are in fact (ecologically) rational, when taking into account environmental variables (Gigerenzer, 2018). For instance, the gambler’s fallacy can represent a probabilistically valid intuition under certain circumstances. Let’s consider that a coin is tossed four times in a row. What sequence of three outcomes is more likely to be encountered: head-head-tail or head-head-head? In fact, if we consider a particular sequence of outcome with length $k = 3$ and a total number of tosses $n = 4$, the head-head-tail is more likely (Fig. 1.4).

| | | | | | | | | | | | | | | | |
|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| H | H | H | H | H | H | H | H | T | T | T | T | T | T | T | T |
| H | H | H | H | T | T | T | T | H | H | H | H | T | T | T | T |
| H | H | T | T | H | H | T | T | H | H | T | T | H | H | T | T |
| H | T | H | T | H | T | H | T | H | T | H | T | H | T | H | T |
| ✓ | ✓+ | + | + | - | - | - | - | ✓ | + | - | - | - | - | - | - |

Figure 1.4: Possible outcome sequences when a coin is tossed four times. A cross is written in the last row when the sequence head-head-tail appears. A check mark is written in the last row when the sequence head-head-head appears. The intuition that head-head-tail is more probable (intuition often seen as violating the statistical normative model) is in fact valid if we consider four tosses, and sequences of length three. The table is taken from [Gigerenzer, 2018](#).

Among 16 possible sequences, HHT is encountered four times ($4/16=.25$) whereas HHH is encountered 3 times ($3/16=.1875$). Similarly, it can be shown that HHT is more likely to appear first under these parameters ([Hahn and Warren, 2009](#)). The only condition for this observation to hold is that $k < n$ and that $n < \infty$. As $n < \infty$ is a reasonable assumption considering humans are mortal, and $k < n$ is valid as long as judgment is required on a smaller amount of observations than the overall sample (think of working memory span limitations for instance), it can be assumed that the cognitive process underlying the gambler’s fallacy is in fact frequently ecologically rational. Several classical biases when seen through the lens of ecological rationality, may appear as intelligent inferences rather than logical or statistical errors (e.g. on the conjunction fallacy, see [Hertwig et al., 2008](#)).

Thus, against the ‘vagueness’ criticized in the heuristics-and-biases program, the fast-and-frugal-heuristic program argues for falsifying formal models of heuristic by prediction, not by data fitting *a priori* ([Gigerenzer and Gaissmaier, 2011](#)). Therefore, model competition is favored ([Berg et al., 2010](#)).

Moreover, [Gigerenzer \(2018\)](#) argues that ‘statements about the rationality of judgments need to be qualified with respect to ecological conditions’, meaning that formal models of heuristic need to include parameters and rules that implements cognitive processes in relation to the environment. In accordance with Simon’s bounded rationality ([Simon, 1955](#)), these heuristics must maximize an accuracy-effort trade-off. A heuristic is thus defined as:

[...] a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods. (Gigerenzer and Gaissmaier, 2011).

A way to make heuristics comparable across environments is to base them on common building blocks:

- **Search rules:** specify in what direction search extends in the search space.
- **Stopping rules:** specify when search is stopped.
- **Decision rules:** specify how the final decision is reached.

Also, in decision under uncertainty where information is scarce and environments possibly complex, heuristics relying on these building blocks can make sense of decisions while reducing environmental complexity and computational cost. In these situations, relying only on the best cue available may be a reasonable alternative. A class of heuristics known as “one-reason decision making”, among which the *take-the-best* heuristic is the most notorious, makes this assumption (Gigerenzer and Goldstein, 1999).

The take-the-best heuristic is a model of inference between two alternatives, evaluating one criterion and based on binary cue values retrieved from memory (Fig. 1.5). Consider the task to infer between alternative A or B, which one has a higher value on a numerical criterion. Let’s say you have to decide whether the German city of Cologne has a larger population than another city, for instance Stuttgart. Let’s denote two cue vectors X_A and X_B , one for each alternative. These vectors are composed of binary cues noted x_A^i and x_B^i , where i is the cue identifier. A cue x_i can refer to a question such as “is this city a state capital?” or “does this city as a soccer team playing in national league?”, by taking a value of 0 or 1. Using the take-the-best heuristic, one individual will therefore:

- Search through cue vectors X_A and X_B , and select a valid cue (cues are ranked by validity). Cues are then compared. While both cues take the value 0 (— in figure 1.5) or 1 (++ in figure 1.5) the search continues.

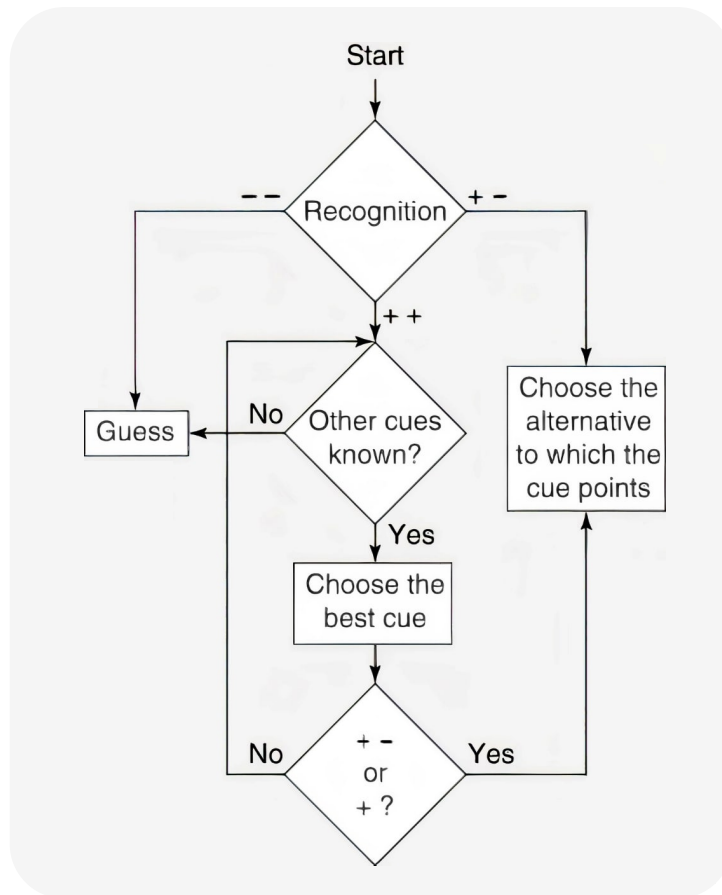


Figure 1.5: Take the best heuristic. The heuristic consists in searching through relevant cues that discriminate between two alternatives, until one cue is found higher on a certain criterion (+−). While both cues have an equal value (++ , −−) the search continues. The figure is adapted from [Maldonato et al., 2011](#)

- Stop the first time a cue x_i discriminates between the options, that is to say if $x_A^i > x_B^i$ or if $x_A^i < x_B^i$ (+− in figure 1.5).
- Ultimately choose the option with the larger value, as-if it has also a larger value on the criterion (here population).

A cue validity is assessed with the following computation:

$$v = \frac{C}{(C + W)}$$

where v is the validity criterion according to which cues are ordered, C is the number of correct inferences when a cue discriminates, and W is the number of wrong inferences.

In some decision contexts, the take-the-best heuristic has been found to account better for subjects' decisions than linear models (Czerlinski et al., 1999), or Bayesian inferences (Bröder and Schiffer, 2003; Dieckmann and Rieskamp, 2007).

Simple heuristics sharing these characteristic building blocks, also have found empirical support. For instance, the *priority heuristic* has been found to explain several deviations from EUT (e.g. Allais' paradox, certainty effect) without implementing subjective value-computation (Brandstätter et al., 2006).

1.4 Summary

Utility theories are rooted in the will to quantify economic value. The labor theory of value was an attempt by classical economists to explain the process by which a good is assigned an economic value. It suggested that the value of a commodity could be measured objectively by the quantity of labor needed to produce it, such that the value is somehow incorporated in the good in question (Smith, 1776; Ricardo et al., 1835; Marx, 1873). Thus, the labor theory of value proposed a materialist definition of value, where economic value derives from production factors. During the 19th century, the labor theory of value was hegemonic (Dillard, 1945), until the marginalist revolution occurred. Marginalists (Jevons, 2013; Menger, 1871; Walras, 1896) conceived value as subjective, and name it 'utility', i.e. the usefulness or pleasure one can derive from the acquisition of a good or a service. They inherited this hedonistic conception of economic value from utilitarianism (Bentham, 1789; Mill, 1859), a moral theory asserting that doing good consists in maximizing pleasure. However, measuring sensations was a nearly impossible task at the time, and numerous debates surrounded this topic (Box 1.1).

Through the ordinal revolution and revealed preference theory (Samuelson, 1938), utility theory gradually got rid of its psychological entities, respectively by seeing utility as ordinal (relative) and not cardinal (absolute), and by considering its measure should be directly derived from indi-

viduals' choices. This absence of ontological commitment toward utility culminated in Friedman's positivism (Friedman, 1953). On this basis, EUT emerged as an influential axiomatic framework for decision theory. However, debates concerning the appropriate normative definition of utility (Box 1.3) arose when several violations of EUT were observed in human decisions (Allais, 1953; Ellsberg, 1961). It led to the creation of the judgment and decision-making field, which favored the empirical study of utility through descriptive models of economic decisions. Interestingly, psychological methods became central again in this approach.

In value-first models (i.e. models that compute an expected-utility for each isolated options), "experience utility", as an hedonic quality (Bentham, 1789), was distinguishable from "decision utility", i.e. the theoretical weighting of an outcome¹⁰. Prospect theory (Kahneman and Tversky, 1979) belongs to this class of models. Similarly in value-comparison models (i.e. models which decisions are based on the computation of utility differences), regret and disappointment minimization are central (Loomes and Sugden, 1982). In short, psychological entities are invoked again, in order to account for utility. For this class of models, lotteries describing full information (probabilities, outcomes) are used to elicit risk-attitudes, which are subsequently used to build utility functions (Holt and Laury, 2002). This experimental setup is also known as the *description* paradigm.

Moreover, models of utility that account for decisions, are put in competition, sometimes using fMRI methods to adjudicate between models. This is the 'neuroeconomics approach', which suggest that utility is not only a theoretical construct, but is also neurally implemented (Camerer et al., 2005). Hence, utility as a scientific construct, has experienced significant ontological (what is it?) and epistemological (how to study it?) variations over time (Moscati, 2018b).

Finally, alternative research programs were also proposed. For instance, the fast-and-frugal-heuristics program advocates for the use of value-free models, and other norms of 'parsimonious' rationality

¹⁰Kahneman et al., 1997 claim that cardinal and empirical measures of utility are 'back to Bentham'.

(Gigerenzer, 1991, 1996; Gigerenzer and Gaissmaier, 2011; Gigerenzer, 2018). Heuristics are strategies that follow a set of rules in order to maximize an accuracy-effort trade-off. The rationality of these heuristics is ecological, in the sense that it has to be assessed relatively to how a heuristic performs in a particular environment.

Physics did not advance by looking more closely at the jubilation of a falling body, or biology by looking at the nature of vital spirits, and we do not need to try to discover what personalities, states of mind, feelings, traits of character, plans, purposes, intentions, or the other perquisites of autonomous man really are in order to get on with a scientific analysis of behavior.

Burrhus F. Skinner, *Beyond Freedom and Dignity*, 1971

Personally, I am primarily intrigued by the possibility of learning something, from the study of language, that will bring to light inherent properties of the human mind.”

Noam Chomsky, *Language and Mind*, 1968

2

Reinforcement Learning

2.1 Behavioral reinforcement learning

2.1.1 Classical conditioning

In the 1890s, the Russian physiologist Ian Pavlov conducted an experiment on the gastric function of dogs by collecting secretions from their salivary gland (Pavlov and Gantt, 1928)¹. He noticed that dogs tended to salivate before they were actually fed, so he decided to study this anticipated physiological reaction. It turned out that this effect was not confined to a chemistry phenomenon,

¹The behavioral experiments he conducted in the 1890s will finally result in a paper translated in English in 1928.

which piqued his interest and led him to conduct a series of experiments. He varied the stimuli occurring after the food was presented, hoping that it will elicit a response and aiming at creating new causal associations. The procedure consisted of delivering food to dogs (an unconditioned stimulus, US), following a tone presentation (a conditioned stimulus, CS) (Fig. 2.1). The conditioned stimulus alone could not elicit salivation (an unconditioned response, UR) at first, yet after numerous tone-food (CS-US) presentations, the dog's salivation could be elicited via both the conditioned and unconditioned stimuli. In this way he discovered the basic principles under the acquisition of conditional response (CR) - that is, reflex responses, such as salivation, that could be reproduced through the association with a novel stimulus. Pavlov saw this phenomenon, which was to be known as 'classical conditioning', as the basis of learning.

2.1.2 Operant conditioning

During the same decade, the 'associative learning' paradigm was concomitantly developed by Edward L. Thorndike (1898). His thesis, *Animal Intelligence: An Experimental Study of the Associative Processes in Animals*, is based on a series of experiments in which cats locked in a box must uncover the mechanism that allows them to break free and access food (Fig. 2.1). The cats move around the box with no apparent purpose and then discovers the action (pulling a rope or pressing a lever for instance) that provides the solution. After several attempts they defeat the 'puzzle' faster and faster. Contrasting to Pavlov setup, Thorndike's box conditions the receipt of the reward (the food) on a behavioral response (enabling a mechanism). This associative learning process where the strength of a behavior is modified by reinforcement was characterized by Thorndike as 'instrumental learning'. This was driven by the *Law of Effect*: responses resulting in a satisfying effect in a specific situation become more likely to occur in that situation, while responses that produce a discomforting become less likely to occur again in the same situation. Said differently, learning occurs via a trial-and-error process, where consequences affect future actions differently. A positive reinforcer will strengthen the link with a behavior, while a negative one will weaken the association.

Thus, there is an important theoretical and experimental distinction between Pavlovian and operant conditioning. In the former the animal only observes the relationships between events in the world, whereas in the latter it also has some control over their occurrence.

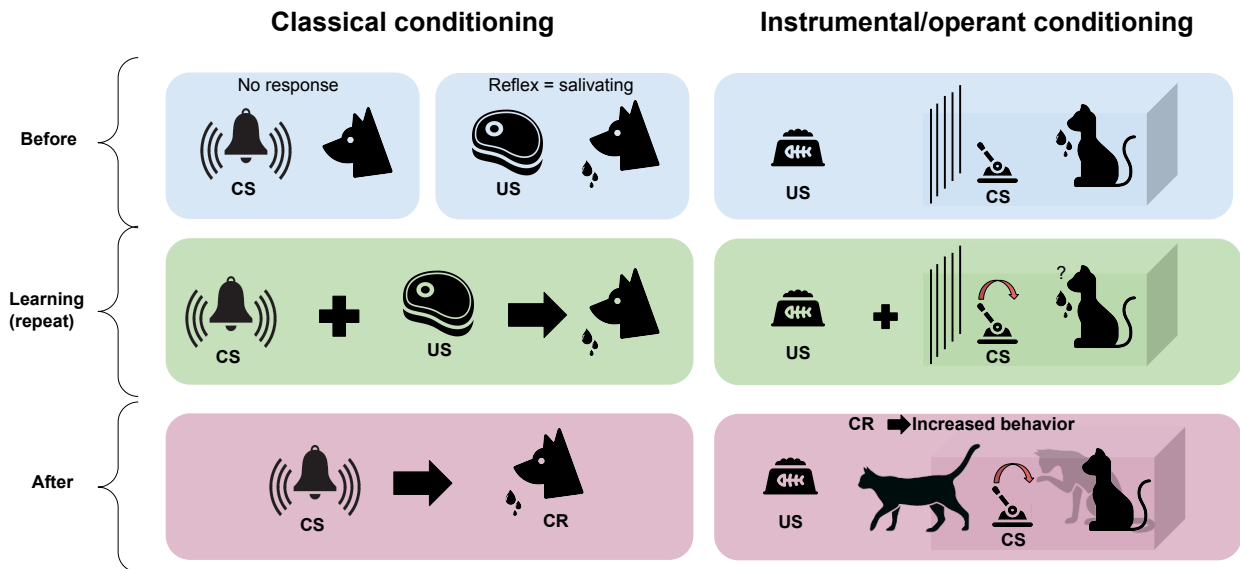


Figure 2.1: Classical and operant conditioning. In classical conditioning, a conditioned stimulus (CS) is associated to an unconditioned stimulus (US) through learning, and elicits a conditioned response (CR). For instance, a steak (US) when presented to a dog, elicits salivation, when a bell (CS) does not. During the learning phase the steak and the bell are presented together. After a sufficient number of repetitions, the bell (CS) is associated to the steak (US) and elicits salivation (CR) when presented alone to the dog. In operant conditioning, the process is similar, but the CS requires an action. As the CS and US association is reinforced, the behavior is reinforced. For instance, a cat is locked in a cage, with a lever (CS) inside and food (US) outside.. The cat must learn to pull the lever (CR) in order to gain access to the food (US). This behavior is thus reinforced and increases.

2.1.3 Behaviorism

A proto-behaviorist (Malone, 2014) approach can be seen in the work of Watson (1920), and notably the *little Albert experiment*, where he showed empirical evidence of classical conditioning in humans. In short, Watson followed the procedures given by Pavlov’s experiments. He first exposed a child, Albert, to a series of stimuli such as rats or rabbits, but also non-animal objects stimuli such as cotton or wool. During those baseline tests, Albert expressed no fear. Thereafter, he tried to elicit an emotional reaction in Albert, by playing loud sounds when Albert was interacting with those stimuli. Watson observed that he created an (aversive) association with those

stimuli, which had been at first US, and become CS, provoking a CR, i.e. a negative emotional reaction (cries, distress) from the little Albert.

Three decades after Pavlov and after acknowledging his work, Burrhus F. Skinner (1938) further extended associative learning theory, and in this way officially establish one the most influential experimental psychology research program in the 20th century: behaviorism (Box 2.1).

One of his primary contributions was the extension of the operant conditioning paradigm, by designing the Skinner box, also called the operant conditioning chamber (Fig. 2.2). The box contained an electrified grid, a food dispenser, a speaker, and a cue light, as well as two levers. The experimenter can use this setup to explore classical (speaker, lights) and operant (levers) conditioning in a variety of species, most commonly rats. The Skinner box's allows for the investigation of several forms of learning:

- **Positive reinforcement:** The rodent presses the lever, and obtains food, resulting in an increase in of the operant behavior due to the association with a reward.
- **Negative reinforcement:** The rodent receives electric shocks, presses the lever, which stops the shocks. The operant behavior frequency is consequently increased in order to avoid the shocks.
- **Positive punishment:** The rodent presses the lever, receives an electric shock, leading to a decrease of the operant behavior by association with a punishment.
- **Negative punishment:** The rodent receives food, presses the lever, the food disappears. The operant behavior frequency is reduced by associating it with the removal of the reward.

By means of this conditioning box, Skinner established the occurrence probability of an action as the most adequate measure of associative strength (Skinner, 1938). The response rate progressively became the main dependent variable considered in the study of operant learning.

This led to the formulation of the *Matching Law* (Herrnstein, 1961), which states that different rates of reinforcement imply different rates of responses. The authorship of this law belongs to Herrnstein, who conducted an experiment on pigeons using Skinner's box. Pigeons were presented

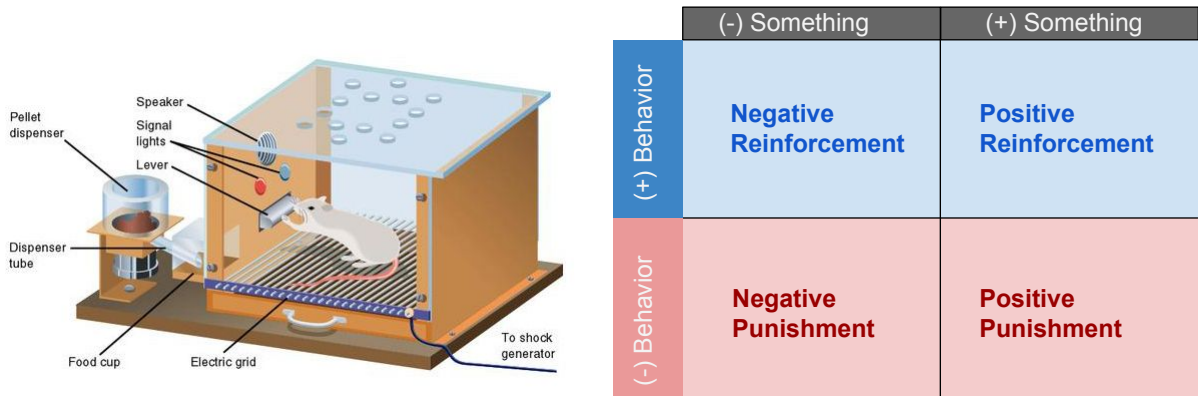


Figure 2.2: Skinner’s box. The box was designed to study various kind of conditioning, notably classical conditioning (with passive stimuli such as tones emitted via a speaker) or operant conditioning (with a lever that enables actions). A system allowed to distribute food, in order to study appetitive conditioning. In addition, an electrical grid allowed to apply punishments, and therefore study aversive conditioning.

with two buttons (A and B), each of them associated to different rates of food reward. Pigeons’ preference went toward the button A, that is the button associated to the greatest food frequency. Interestingly, Herrnstein observed that the ratio of the reinforcement was equivalent to the ratio of responses among the two alternatives. Hence, the *Matching Law* is formally expressed as follow:

$$\frac{R_A}{R_A + R_B} = \frac{Rf_A}{Rf_A + Rf_B}$$

With R_A and R_B the rate of responses that have different rates of reinforcement Rf_A and Rf_B , the matching law holds that the relative response rate matches the relative reinforcement rate.

Box 2.1: Measurements in Psychology and Behaviorism

In parallel to the psychophysics debate mentioned in chapter 1, several sub-disciplines of psychology were committed to measurements. Differential psychology for instance, through scientists such as Binet (1907), was interested in measuring psychological variables (e.g. intelligence, memory) through a series of tests (e.g. questionnaires) that aimed at direct measurements. Others such as Ebbinghaus (1913) took an associationist (i.e. the idea that mental processes operate by the association of one mental state with other states) approach for studying memory, by building learning (or performance) curves. He tested the memorization of nonsense syllables, and concluded that performance decreased depending on several factors (e.g. difficulty of the learned material, physiological variables such as stress, etc.). Behaviorism emerged concurrently, and proposed a radically different program for psychological science, that rested (in its most radical form) on three claims (Graham, 2019):

- Ontological reduction: Psychology is not the science of the inner mind. Psychological phenomenon should be ontologically reduced to behavioral phenomenon.
- Causal reduction: Psychology as a science should not make appeal to mental entities or events. Causes of behavior are external (physical events in the environment) not internal (the mind). Therefore models of psychology should be concerned with inputs (stimuli) and outputs (behavior).
- Epistemic reduction: Mental terms and concept should be eliminated, or if possible translated to behavioral concepts.

Supporting one of these three claims is to be considered as a behaviorist. Within behaviorism, several schools of thought can be distinguished, such as the mathematical approach of Clark Hull, or the radical behaviorism of Skinner. Hull (1932) looked at the performance of rats learning simple tasks such as discrimination the correct arm in a T-maze, developed learning rules of the form $V(t + 1) = \alpha(1 - V(t))$, where V is response strength and α is a learning rate (Staddon and Niv, 2008). Skinner for its part, likely accepted the three above mentioned claims (Graham, 2019). He saw value only in predicting behavior, not in modeling cognitive processes^a. He rejected the Hullian approach, and conceived a set up (the skinner box) where animals are treated "much like physiological preparations" (Skinner, 1956; Guttman, 1977; Staddon and Niv, 2008).

^aHe might, in this regard, be reconciled with Friedman's positivism and instrumentalism (Box 1.2)

2.1.4 Blocking effect

Studying rats' learning behaviors, psychologist Leon J. Kamin reported a striking phenomenon (Kamin, 1967a,b). He showed that a prior CR training to a first stimulus (CS_A) undermines the acquisition of a second CR to second stimulus (CS_B) if presented together as compound stimulus (CS_{AB}). Experimentally translated, the operant behavior was pressing a button (CR) for a reward (US) by food restricted rats. The CS were light (CS_A) and electric shock (CS_B). After several

trials, the paired association $CS_A - US$ is successfully learned. For the animal, a light predicts the occurrence of the reinforcer (the food), and a CR is acquired, which consists in pressing the button. Thereafter lights and shocks (CS_{AB}) are associated to the occurrence of the food. However, thereafter when shocks are used alone (CS_B), the response is diminished². In Kamin's reasoning, this *blocking effect* suggested that traditional theories of classical conditioning were incomplete, given that the latter assumed that contiguity of a CS with a US is a sufficient condition for the establishment of a CR. Kamin consequently argues that this phenomenon reflects an higher order cognitive process, such as attention, predictability or surprise. Indeed, assuming contiguity is sufficient, why are associations not formed with other stimuli present in the environment, like the experimenter for instance? Some mental process must have selected specific stimuli for learning to occur. Kamin's call for cognitivist explanations were thus conflicting with the input-output model endorsed by behaviorists. As Moore and Schmajuk claimed:

The strength of Kamin's evidence from his blocking experiments fueled the then nascent cognitive perspective, which in the ensuing decades became a dominant feature of modern learning theory and computational models of classical conditioning. (Moore and Schmajuk, 2008).

2.2 Computational reinforcement learning

2.2.1 Basic principles

Progressively and based on the work of early behaviorists, Reinforcement learning (RL) as a formalized learning framework arose (see Box 2.2). Instead of being explicitly taught a goal, an RL agent learns from the consequences of its own actions. It selects actions on the basis of its past experiences (exploitation) and also by exploring new choices (exploration), or said otherwise, via

²A first control consists in modifying the learning architecture, by skipping the prior conditioning $CS_A - US$ and presenting the compound stimulus CS_{AB} directly, which effectively results in a CR (pressing the button for food) when only CS_B is presented. A second one consists in first learning the $CS_{AB} - US$ association, then learning the $CS_A - US$ association, which again do not result in blocking the CR for CS_B alone.

a trial-and-error process. The training signal that the agent processes is a numerical reward encoding the success of an action's outcome. In addition, the agent seeks to maximize the expected accumulated reward over time.

The agent integrates the training signal by means of a *learning rule*, such as the *delta rule*. The balance between exploration and exploitation is defined via a *policy* (i.e. how to select an action, learning rule put aside), such as the *softmax policy* or the *epsilon-greedy algorithm*.

The RL framework is frequently thought as a Markov Decision Process (Howard, 1960). A Markov decision process is a discrete stochastic process. At each step, the process is in some state s and the agent chooses an action a . The probability that the process arrives at the state s' is determined by the chosen action. More precisely, it is described by the state transition function $T(s, a, s')$. Thus, the realization of the state s' depends on the current state s and the selected action a . We then say that the process satisfies the Markov property. When the process moves from state s to state s' , the agent receives a reward $R(s, a, s')$.

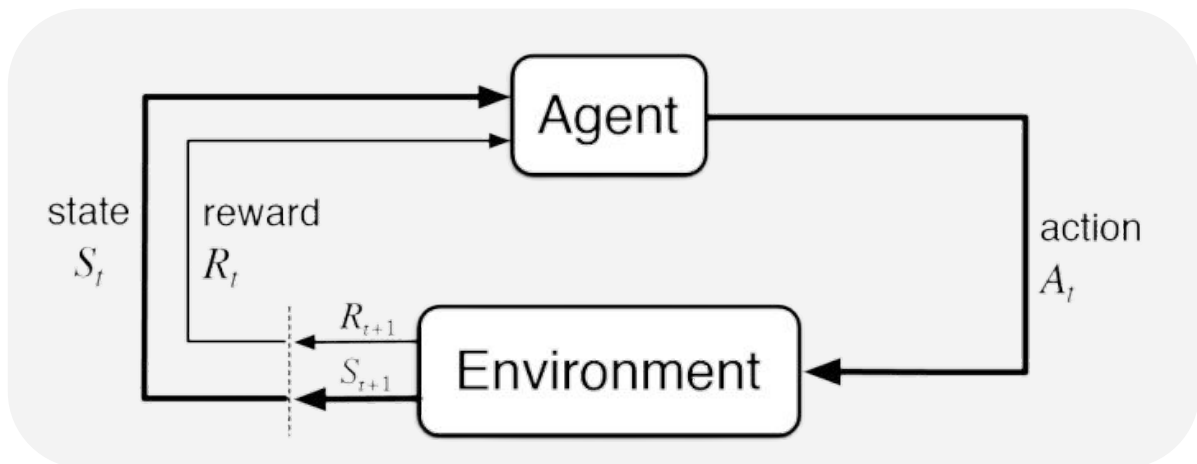


Figure 2.3: The basic reinforcement learning components. At time step t and in state S_t , an agent interacts with the environment by selecting an action A_t . Thereafter, it enters a new time step $t + 1$ and state S_{t+1} while obtaining a reward R_{t+1} .

Thus, the RL framework can be defined by:

- a set of states S .

- a set of actions A .
- $P(s, a, s')$; the probability of transition from state s to state s' under action a .
- $R(s, a, s')$; the immediate reward after transition from s to s' with action a .

2.2.2 Prediction Problem and Control Problem

RL is used to address two kinds of problems (Woergoetter and Porr, 2008):

- **Prediction problem:** the agent learns the value function for the policy followed. When the algorithm converges, it possesses a value function that must encode the maximum expected reward for every visited state.
- **Control problem:** the agent seeks to find a policy which maximizes the expected reward when traveling through state space (i.e. by interacting with the environment). As the agent has a control over the state sequence, it must learn how to travel optimally depending on the feedback it receives in different states. The *control problem* is more demanding than the *prediction problem*, as it implies to solve the prediction problem as well.

Historically, formal models of classical conditioning coming from animal learning are more concerned with the *prediction problem* (Balkenius et al., 1998). Animals typically remain passive and no specific actions or interaction with the environment are required from them. Among this class of models, the Rescorla-Wagner model (Rescorla, 1972) and the Temporal-Difference model (Sutton and Barto, 1981; Sutton, 1988) are probably the most notorious.

The control problem on the other hand is present in operant conditioning (as animal actions possibly affect the environment), and is notably addressed by the Actor-Critic architecture, which is essentially derived from the Temporal-Difference learning method. Furthermore, another model specifically designed to reach optimal control over the environment is the Q-learning model (Watkins, 1989).

2.2.3 Rescorla-Wagner Model

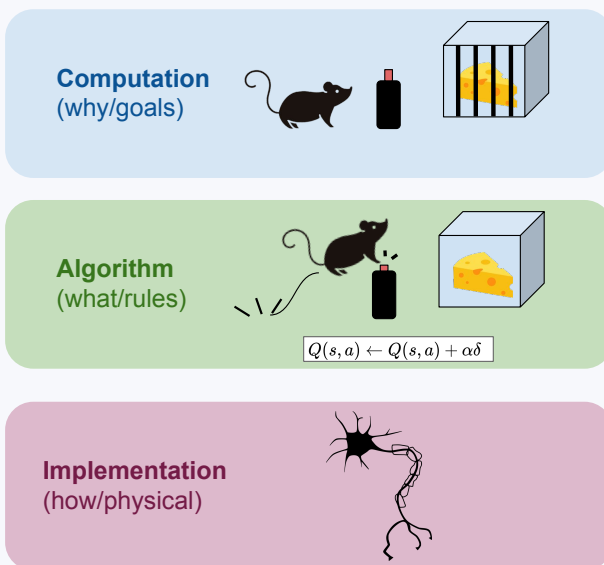
In 1972, Yale psychologists Robert A. Rescorla and Allan R. Wagner present a model of classical conditioning (Rescorla, 1972), that provides a formal explanation for the *blocking effect*. An assumption fundamental to the model is that the total amount of conditioning the stimuli may obtain from the reinforcer is limited. This finite quantity is shared among the stimuli of which the compound is composed. Another assumption is that the salience of the stimuli modulates the associative strength. Formally, for a pair of stimuli A and B, the associative strength is thus incremented at each trial as follows:

$$\begin{aligned}\Delta V_A &= \alpha_A \beta_A (\lambda - V_{tot}), \\ \text{and } \Delta V_B &= \alpha_B \beta_B (\lambda - V_{tot}), \\ \text{where } V_{tot} &= V_A + V_B\end{aligned}$$

where ΔV_A and ΔV_B are respectively the gradient strength of the association of stimulus A and B with the reinforcer. V_{tot} is the total associative strength of the compound. The salience of each stimulus is implemented by α , while λ is the asymptote of conditioning (i.e. the maximum amount of conditioning). Finally, β implements the learning rate of a particular US (the idea being that different US might be perceived as more or less attractive, and therefore might reinforce with different intensity). Three elements thus predict the amount of conditioning: (i) the salience of the CS (ii) the learning rate parameter associated to a US (α) (iii) the difference between the asymptote (λ) and the associative strength of all the cues present in the environment. V increases when the difference is positive, while it decreases if it is negative. This error-correction mechanism is crucial, as it implements the psychological 'surprise', which ultimately allows overwrite previous information. In the next years, the Rescorla-Wagner model gained influence, due to its ability to explain a wide range of behavioral features observed in conditioning tasks, in a parsimonious manner (Miller et al., 1995; Siegel and Allan, 1996). For instance, the model predicts Kamin's blocking effect, merely because after CS_A is learned, V_{tot} will be near the asymptote λ , which

results in preventing new associations. Despite this explanatory power, several limitations will foster the emergence of alternative models. Importantly, the model is agnostic regarding higher-order structure of the environment, and thus can't implement the relation between actions and different states for instance. It is thus restricted to *prediction problems* and cannot address the *control problem*. Furthermore, its approach to the *prediction problem* is also limited, as it can only learn from immediate outcomes, and not sequences of trials (Sutton and Barto, 1981; Gaffan, 1989).

Box 2.2: Marr's levels



RL gained massive popularity over the last thirty years. For instance, the occurrence of the 'reinforcement learning' term was multiplied by 60 among nature journals (Niv and Langdon, 2016). Arguably, this success is due to the idea that RL is a computational neuroscience framework that encompasses all three levels of Marr (Marr and Poggio, 1976). Let's imagine a rat put in experimental conditions, and that aims to eat a cheese locked in a cage. At the *computational level*, a set of goals and assumptions are defined, for instance maximizing expected reward, i.e. getting as most food as possible.

At the *algorithmic level*, the rules that guide behavior in order to achieve the goals are defined. This is typically what RL models formalize, i.e., learning rules and policies that aim at fulfilling predefined goals. Here, the rat must press a button (i.e. select an action) in a particular context (i.e., a state), to obtain the reward, that will act as a reinforcer of the behavior. Lastly, the *implementational level* describes how this behavior is neurally and physically performed. Typically, models' variables are linked to the activity of neural substrates, via electrophysiological methods or fMRI. For instance, the prediction-error δ , that implements the difference between the expected and obtained outcome, is notoriously correlated to dopaminergic activity in the basal ganglia (e.g. Schultz et al., 1997, Pessiglione et al., 2006).

2.2.4 Delta rule for Neural Nets

In parallel to the development of the conditioning paradigm, the nascent machine learning field was developing. In 1943, the neurophysiologist Warren McCulloch and the mathematician Walter Pitts published an article describing the functioning of neurons by representing them as electrical

circuits (McCulloch and Pitts, 1943). This representation was to become the theoretical basis for neural networks and other standard connectionist models. Later, Frank Rosenblatt (1958) invented the perceptron, which consists in formal neurons where synaptic weights are updated through a learning rule. ADALINE, an extension of the perceptron was proposed by Widrow and Hoff (1960). More specifically, ADALINE is a single layer neural network with multiple inputs neurons which generates one output. The inputs are connected to the output neuron through weighted synapses. Formally, the output y is computed as follows:

$$y = \sum_{i=1}^n x_i w_i$$

with x being the input vector and w the vector of synaptic weights. For each iteration in the convergence process, the weights are updated via the following mechanism:

$$w \leftarrow w + \eta(d - y)x$$

with w being the synaptic weight, η the learning rate, d the desired output, and o the actual output of the network. Depending on the class of problem (single layer networks are known for their inability to treat data that is not linearly separable), the network will eventually converge such that the output y gets closer to d . This kind of learning mechanism is called 'supervised', as there is a prior knowledge of the desired final output.

2.2.5 Temporal difference learning

In the 1980s and the early 1990s, Sutton and Barto proposed the Temporal Difference (TD) learning algorithm (Sutton and Barto, 1987; Sutton, 1988; Sutton and Barto, 1990). Inspired by the correspondence they observed between the Rescorla-Wagner and ADALINE learning rules (Sutton and Barto, 1987), they aimed at constructing a new model of classical conditioning, that would borrow elements from optimal control (machine learning) and animal learning. More precisely, they pointed out that in the ADALINE model, the reinforcement signal ($d - y$) implied by the delta rule

presented surprising similarities with the Rescorla-Wagner learning rule (Sutton and Barto, 1987, 1981). The equivalent term in Rescorla-Wagner is thought to implement the psychological 'surprise' when facing a stimulus. In the ADALINE model, this 'surprise' phenomenon translates as the the difference between the expected outcome and the obtained outcome. According to Sutton and Barto, those two models thus belong to the class of prediction-learning methods. However, they also note this class of models lacks a crucial dimension in their implementation, temporality. When describing classical conditioning, the Rescorla-Wagner model specifies changes in associate strength at a trial level. In contrast, TD methods allow for 'real-time' models (as Sutton and Barto named them), in the sense that they are able to update associative strength from sequences of trials:

Whereas conventional prediction-learning methods are driven by the error between predicted and actual outcomes, TD methods are similarly driven by the error or difference between temporally successive predictions; with them, learning occurs whenever there is a change in prediction over time. For example, suppose a weatherman attempts to predict on each day of the week whether it will rain on the following Saturday. The conventional approach is to compare each prediction to the actual outcome whether or not it does rain on Saturday. A TD approach, on the other hand, is to compare each day's prediction with that made on the following day. If a 50% chance of rain is predicted on Monday, and a 75% chance on Tuesday, then a TD method increases predictions for days similar to Monday, whereas a conventional method might either increase or decrease them depending on Saturday's actual outcome. (Sutton, 1988)

Let us consider an agent traveling to a sequence of states and rewards, during T timesteps.

$$s_t, r_{t+1}, s_{t+1}, r_{t+2}, \dots, r_T, s_T.$$

Let R_t be the sum of all the rewards obtained at the current state:

$$R_t = r_{t+1} + \gamma r_{t+2} + \dots + \gamma^{T-1} r_T$$

where r_{t+1} is the immediate reward and $\gamma \in [0, 1]$ is the discount factor, which is weighted to give more importance to recent benefits while discounting (i.e. disregarding) future outcomes more strongly. If the complete return, i.e., the cumulative future reward R_t expected from this state s_t , is dependent on the value of the current state $V(s_t)$, then we can estimate the value of a state using a delta rule, which implements an error-correction signal:

$$V(s_t) \leftarrow V(s_t) + \alpha \cdot (R_t - V(s_t))$$

with $\alpha \in [0, 1]$ being a learning rate parameter, which determines the extent to which this signal will override previous information. When $\alpha = 0$, $V(s_t)$ remains identical because the agent ignores the reinforcement signal. If $\alpha = 1$ the most recent information completely overwrites previous one. Similarly to the Rescorla-Wagner and likewise ADALINE model, the term $R_t - V(s_t)$ is a reward prediction error (Sutton and Barto, 1981), i.e., the difference between the complete return (obtained reward) and the predicted one (expected reward). When $V(s_t)$ and R_t are equals, the agent perfectly predicts the complete return value and the reward prediction error will be zero and hence the algorithm will converge. In the TD(0) algorithm described by Sutton and Barto, instead of using the accumulated sum of discounted rewards R_t , we only look at the immediate reward r_{t+1} , plus the discount of the estimated value of only one time step ahead $V(s_{t+1})$:

$$V(s_t) \leftarrow V(s_t) + \alpha \cdot (r_{t+1} + \gamma V(s_{t+1}) - V(s_t))$$

To imitate an immediate reward learning (such as in the Rescorla-Wagner model), we can consider the $TD(s = 0)$ case, or in other words, the case where it makes estimates from estimates, instead of estimates from sequences of trials. In this way, the TD-error δ is defined:

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$$

The prediction error signal can be used to reinforce actions leading to desirable states of the environment and discard those leading to worse states. Without considering any other previously

visited states, we assign a new state value to one state by performing:

$$V(s_t) \leftarrow V(s_t) + \alpha \cdot \delta_t$$

TD allows predictions over successive time steps to drive the learning process. The prediction at any given time step is updated, such that at the next time step, the distance between the previous prediction and current prediction will be reduced. In other words, it is a particular class of supervised learning processes, where the training signal for a prediction is derived from future predictions, instead of being computed based on an immediate outcome. This property provides TD models with certain advantages over trial-level models. First, TD allows to account for the frequency of stimuli presentation within a trial, whereas the Rescorla-Wagner for instance cannot. Indeed the inter-stimulus interval between US and CS is known to have strong effect when learning associations (Odling-Smee, 1975). Second, these models, because they are in real-time, are more mechanistic and therefore more amenable to speculation about their physical implementation, in particular in the light of electrophysiological data (Suri and Schultz, 2001). Although it was initially designed to solve the *prediction problem* of classical conditioning, it was later extended to model instrumental conditioning. Thus, it was able to cope with the *control problem*, notably using an actor-critic architecture.

2.2.6 Actor-critic

Actor-critic methods are TD methods that have a separate memory structure, in order to represent the policy independently of the value function (Fig. 2.4). One of their aims is to deal with the *control problem* (Williams, 1992; Sutton et al., 2000).

The actor implements the policy and is used to select actions. The critic is synonymous to the value-function, and is called critic because it adjusts and guides the choices made by the actor. Both components are only informed by the current state. For this reason, they learn policies directly, that is, without calculating option values. The critic therefore learns from the prediction-

error signal and informs the actor on whether to maintain a policy or not. In contrast with the standard TD-learning model, where the action was left unspecified, the actor-critic model allows for action-selection, allowing to describe operant conditioning.

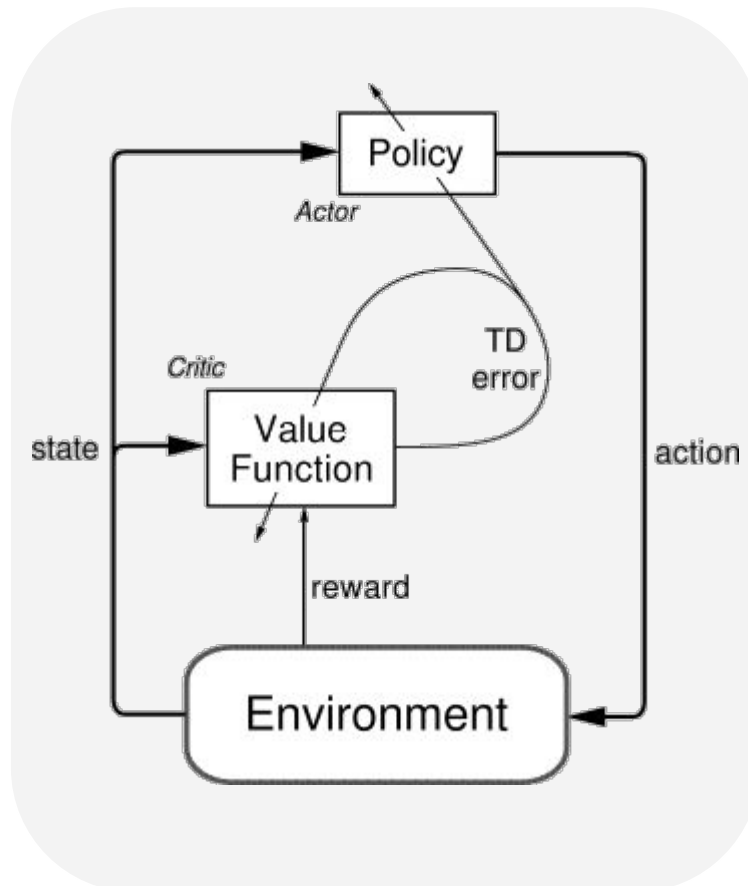


Figure 2.4: Standard actor-critic architecture. Actor-critic methods are temporal-difference learning methods which rely on an architecture based on two main components. The actor is the learning component, while the actor is the decision component. The actor, given the state-value function, selects an action which maximizes expected reward. The obtained reward is used to compute the temporal-difference error, which is both fed to the critic and the actor, in order to respectively update the state-value function and guide future policies.

As standard TD models, the actor component learns from the following prediction-error δ :

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$$

where V is the state-value function implemented by the critic. The actor is also informed about the state, and can select the appropriate action. The specific architecture of the actor-critic, implies

that the actor directly implements the TD error, in order to adjust the weighting of the selected action, with the following update rule:

$$p(a_t, s_t) \leftarrow p(a_t, s_t) + \beta \delta$$

where $p(a_t, s_t)$ is the probability of choosing the action a a time t , β determines to what extent the prediction-error overrides the initial probability.

A desirable feature of this dual architecture, and especially in the machine learning field, is the reduction of the computational complexity. Learning a unique state-value function, instead of a value for each action-state couple (action-state value function), reduces the risk of combinatorial explosion, when the number of states and actions exponentially grows. This is known as the curse of dimensionality, which happens in high-dimensional spaces (Sutton and Barto, 2018).

Moreover, the separate actor in actor-critic methods makes them good candidates for psychological and biological modeling. This dissociation can be used to impose domain-specific constraints, for instance to build model of the ventral and dorsal striatum, where the former corresponds to the critic and the latter corresponds to the actor. (O’Doherty et al., 2004; Takahashi et al., 2008).

In behavioral research, actor-critic architectures were able to account for the matching law and conditional avoidance (Sakai and Fukai, 2008; Maia, 2010).

2.2.7 Q-Learning

Q-learning was first introduced by Chris Watkins (1989). Q-learning differentiates itself from the above mentioned algorithms in the sense that it seeks to maximize a value function for each state-action pair, stored in a Q-matrix $Q(s, a)$. The "Q" in Q-learning stands for quality, i.e. how much an action will lead to future rewards. For each time step t , the agent chooses an action a_t and receives a reward r_t . Following that choice, it enters a new state s_{t+1} . The agent thus interact with an environment composed by a state vector S , with $s \in S$. For each state, there are available actions denoted $a \in A$. Given a state-action couple (s, a) , there is an underlying

outcome probability distribution, where $P(r_t|s_t, a_t)$, is the probability of obtaining the reward r_t . Therefore, the bellman equation (Bellman, 1956) for the Q-learning rule is of the form:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t))$$

where α is the learning rate that controls the weight of new information, γ is the discount factor, r_t is the reward received when moving from state s_t to state s_{t+1} , and $\max_a Q(s_{t+1})$ is the maximum reward that can be expected in the next state. Q-learning became popular in human behavioral research, notably to model bandit task decisions. It was declined under various forms (by using different learning rates for different types of signal for instance), in order to account for a wide range of phenomena such as optimistic bias (Lefebvre et al., 2017), confirmation bias (Palminteri et al., 2017a), or even to account for social learning (Najar et al., 2020).

2.2.8 Action selection and the exploration-exploitation trade-off

The above learning rules provide a way to obtain estimates of action-state values (Watkins, 1989), or only state values (Sutton and Barto, 1981; Williams, 1992; Sutton et al., 2000). However, forming accurate value estimate requires a policy to sample the action-space efficiently. For this reason, several decision rules have been conceived. They rely on parameters that will adjust the exploration (sample new options) and exploitation (maximize the expected reward) trade-off.

We can distinguish three widely used decision rules:

- Argmax rule (Sutton and Barto, 2018): a function that deterministically selects the action with the highest estimated value. By default, there is no exploration.
- ϵ -greedy rule (Sutton and Barto, 2018): a function that selects either the action with the highest estimated value (with probability $1 - \epsilon$) or a random action (with probability ϵ).
- Softmax rule (Luce, 2012): a function that stochastically selects actions. The probability of an action increases with the relative value difference to other actions available in the action

set. Formally, for a Q-learning model:

$$\pi(a) = \frac{e^{\beta Q(s,a)}}{\sum_{a' \in A} e^{\beta Q(s,a')}}$$

with a' being the alternative actions, A the action set, Q the state-action value function, and β the inverse temperature. The inverse temperature indicates the degree of stochasticity. When $\beta = 1$ values are unchanged, when $\beta = 0$ all actions become equally likely. Finally, when $\beta \rightarrow \infty$, the function becomes deterministic and acts as an argmax rule.

2.2.9 Value-free models

In value-based RL models, agents usually maximize an action-value function. The underlying assumption is that at an algorithmic level (see Box 2.2) individual assign values to isolated options in order to compare them. However, some have argued that the value construct is not necessary to describe plausible decision-making processes, and that a parsimonious principle would be not to invoke value as a component of learning models. Several computational models do not integrate a value function, among them the policy gradient models (e.g. Bennett et al., 2021), some connectionist models (e.g. Suri et al., 2020), or habits' models (e.g. Miller et al., 2019). However, we will only describe the case of policy-gradient methods, because it is more in line with the RL models (and especially actor-critic architectures) seen previously.

Policy gradient methods

Policy gradient methods rely upon optimizing parametrized policies with regards to the expected return. Recently, Bennett et al. (2021) described a model where the agent interacts with the environment by selecting actions according to a parameterized policy (Fig. 2.5), such that:

$$\pi^\theta(a) = \frac{e^{\theta_a}}{\sum_{a' \in A} e^{\theta_{a'}}}$$

where a is the considered action, A is the action set, a' is an alternative, and θ is the set of pa-

rameters that adjust the policy toward a particular option. Hence, policy-gradient algorithms do not store expected rewards. The parameters are most of the time to be thought of as representing actions in terms of desirability.

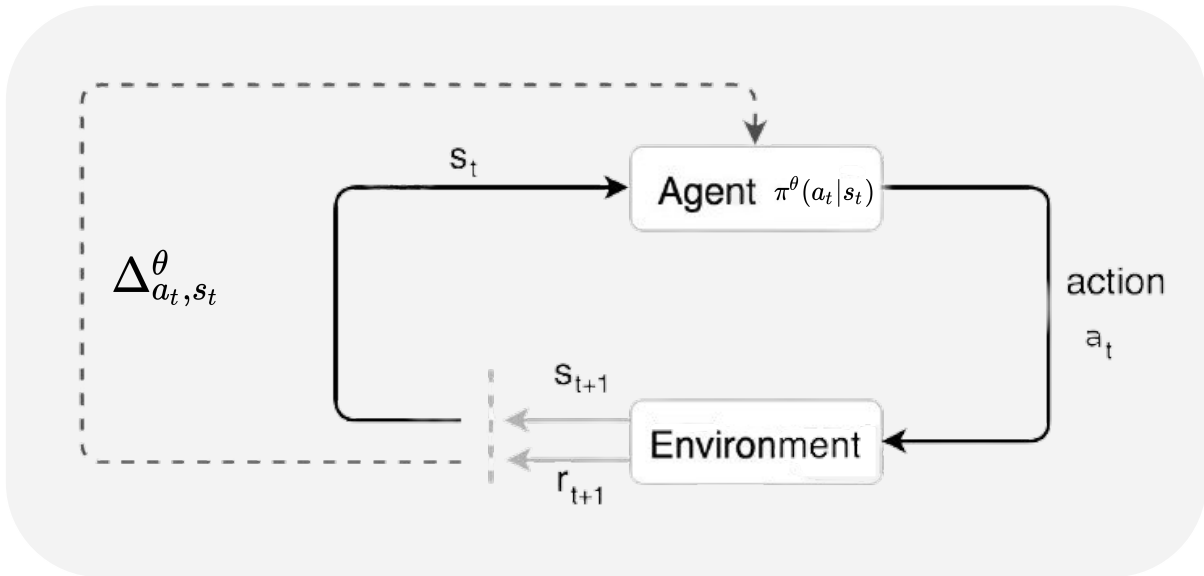


Figure 2.5: Policy gradient learning. A policy-gradient algorithm selects an action a_t given a state s_t according to a parameterized policy π^θ . The set of parameters θ is updated via gradient descent following the obtention of the reward r_{t+1} . In contrast with value-based algorithms, there is no value-function *per se*.

These parameters are directly updated subsequently to the choice, that is after obtaining the reward r_t . Thus, considering the action a and the policy π with parameters θ :

$$\Delta\theta_a = \begin{cases} \alpha \cdot [1 - \pi^\theta(a)] \cdot r_t & \text{if } a \text{ was chosen} \\ -\alpha \cdot \pi^\theta(a) \cdot r_t & \text{otherwise} \end{cases}$$

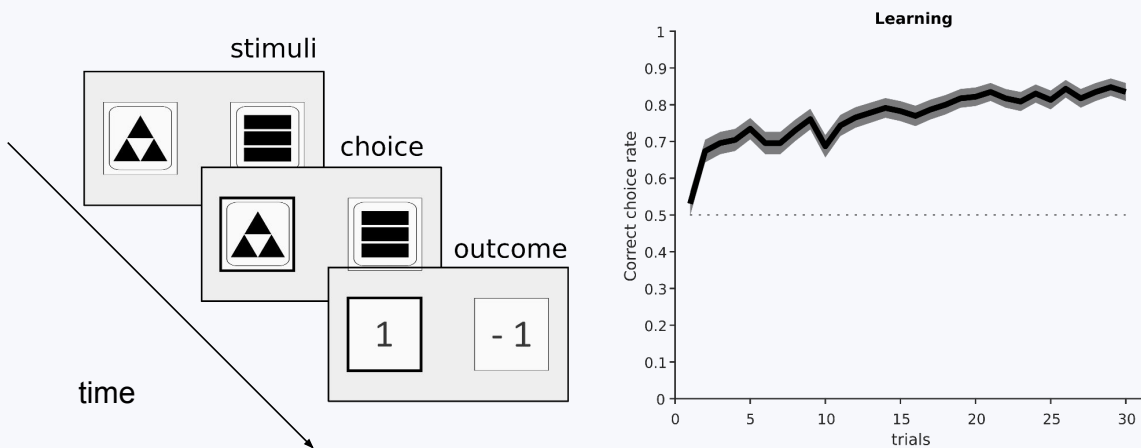
with α being the learning rate controlling the quantity of information that is taken into account to update the parameter set θ . Please note that in this example, we chose to consider an environment where the state is fixed, but some models integrate the state as an input variable (e.g. [Baxter and Bartlett, 1999](#)).

This model shines because of its simplicity: the learning rate constitutes the only free parameter. Furthermore, the absence of a value-function allows the avoidance of the intractability problem that arises from complexity due to continuous states and actions. Indeed, since policy-gradient

methods optimize an overall policy and not action-values, they can be applied without difficulty to a continuous action space.

Box 2.3: The experience paradigm in humans

In most situations, the expected utility of an action is unknown, as decision variables are not described *a priori* (Gigerenzer et al., 2005). Among humans and animals, it is frequent to make choices among options with imperfectly known outcomes, which can only be learned from experience. In neuroscience and experimental psychology, such situations are often translated using the multi-armed bandit task, in which subjects repeatedly choose among options with unknown expected-value, thus negotiating the tension between exploitation and exploration (Sutton and Barto, 2018). Indeed, the multi-armed bandit paradigm has imposed itself as a useful framework to study this trade-off, either theoretically (e.g. Whittle, 1988), or empirically (e.g. Daw et al., 2006).



Thus, abstract cues which convey no information in particular are presented to individuals. Most of the time, individuals are expected to find which options maximizes expected value among a given choice set. The learning curves thus represent the frequency with which subjects choose the most rewarding option, meaning that they progressively reveal the underlying distribution of outcomes for a particular set of options. Furthermore, investigating the generative processes underlying these behaviors requires various methods (Palminteri et al., 2017b; Wilson and Collins, 2019). Manipulating the learning architecture and decision variables (e.g. number of options, quantity of information displayed in the feedback, gain or loss frames, etc) the experimenter can expect to elicit different learning and decision mechanisms. Computational models (e.g. Q-learning; Watkins, 1989) are often put in competition in order to account for these psychological processes. Model comparison and falsification involves finding parameter values that best account for the behavioral data for a given model. The estimated parameters can be used to further explore inter-individual differences, or simulate new datasets under different hypotheses. Additionally, fMRI and electrophysiological methods can allow to corroborate or falsify computational models, by assessing the degree of plausibility of the data under certain model assumptions (Camerer et al., 2004b; Rustichini, 2009).

2.3 Neural reinforcement learning

2.3.1 Value-based decision-making

What are the processes by which our brains adjudicate between two different options? Decision-makers who adhere to axiomatic rationality are assumed to behave as-if they compute and compare utilities (Von Neumann and Morgenstern, 1944). Similarly, behaviorism and RL computational models assume an associative strength between an option, a state, and a reward (Rescorla, 1972; Sutton, 1988; Watkins, 1989), which translates into a value function, allowing measures of subjective expected values. RL and most utility models have in common to suggest that we assign a scalar value to each option, and then select the one with the highest value, i.e. the two-step model of decision-making. However, they were developed within different experimental paradigms (see Box 2.3). But do we physically compute and compare subjective values? Does the as-if computation actually occur in the brain? Economists have been historically reluctant to make ontological commitments regarding value computation (Friedman, 1953). In a similar fashion, important behaviorists have traditionally limited their speculation to a behavioral input-output model, and refrained to suppose mental entities (Skinner, 1956). Positing a valuation stage needs additional empirical evidence, other than choice behavior. Consequently some scholars advocated for a novel research program, grounded on the idea that decisions originate from a neural valuation process. (Camerer et al., 2004b; Rustichini, 2009).

2.3.2 The prediction-error in monkeys

Historically, the quest for finding a neurobiological ground for decision-making models, marked a decisive turning point with Wolfram Schultz. In a seminal paper (Schultz et al., 1997), he showed with an electrophysiological setting that the activity of midbrain dopaminergic neurons encoded the discrepancy between the actual reward and its prediction. As mentioned in the previous chapter, the reward-prediction error (RPE) is a fundamental component of RL models. Functionally,

it implements the update of previously acquired information by integrating new information, i.e., the differences between received and predicted rewards. It hence is conceived as the learning-driving signal. Psychologically, it represents the surprise between one expected outcome and the actual outcome (Rescorla, 1972). Schultz showed that such a surprise, when positive, provokes an increase of activity (better than expected reward), and contrastingly, when negative (worse than expected reward) induces a depression. Also, a correctly predicted reward elicits no response (prediction error of zero) (Fig. 2.6). Translating those results to the conditioning paradigm language, after learning, fruit juice (the unconditioned stimulus; US) was paired with a tone (the conditioned stimulus; CS), such that the tone predicted the dopaminergic response. Interestingly, this pattern of activity perfectly matched the idea that dopaminergic activity encoded the RPE, as it is computationally formulated in TD-learning models.

Thereafter, the subsequent literature used a similar experimental set up, inherited from the pioneering work of Herrnstein (1961). In it, monkeys are typically asked to choose between options by keypress, or touching a screen. Variable amounts of juice allow to manipulate reward magnitude to different options with different color codes (e.g. Platt and Glimcher, 1999). In this paradigm, it has been for instance found that monkeys's choices relied on a reward probability estimated by sampling over the last few trials (Sugrue et al., 2004). Neural activity in response to reward variations has been located in ventral midbrain areas (Fiorillo et al., 2003), and the correspondence between dopaminergic neurons phasic activity and prediction-error has been further established (e.g. Morris et al., 2004). Moreover, modulating this activity has been shown to affect choices (Pessiglione et al., 2006). However, the relevance of the relationship between single cell dopaminergic activity and prediction-error has been questioned. Indeed, some studies found that this relation might only hold for positive prediction-errors (Bayer and Glimcher, 2005). More generally, dopamine signals might serve more than one function (e.g. the vigor in selecting an action relies on dopamine) and are present in various brain areas (Niv and Schoenbaum, 2008).

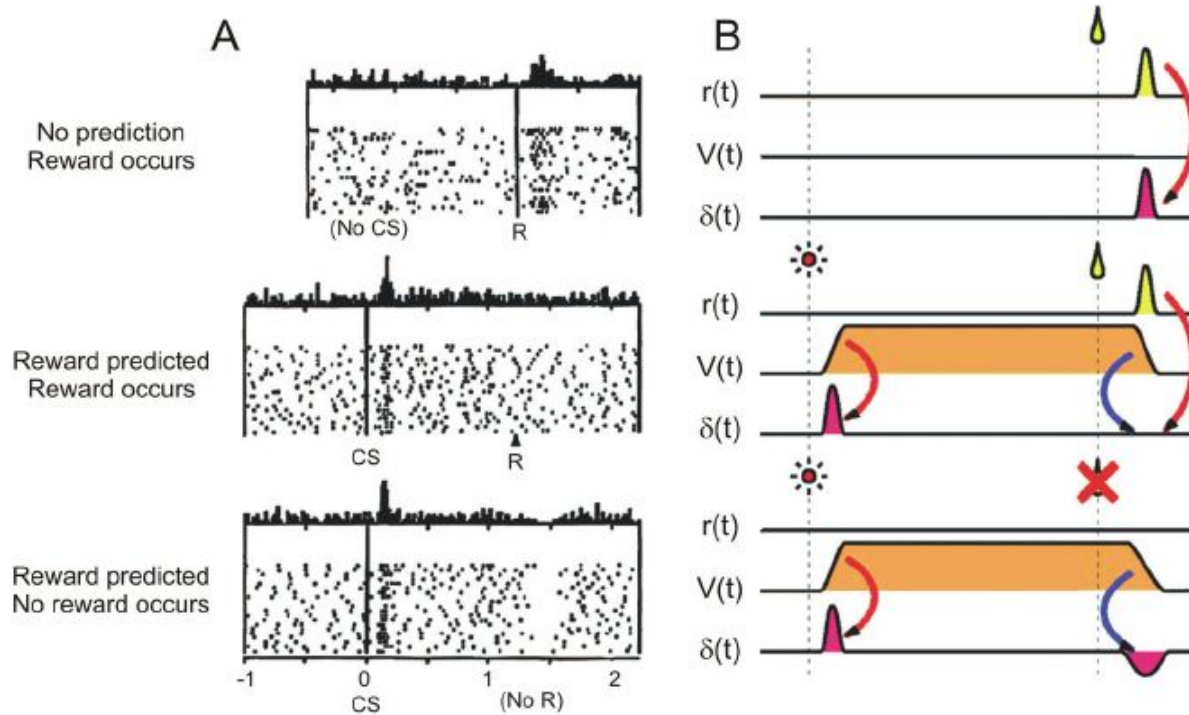


Figure 2.6: Raster plot showing how dopamine neurons encode prediction-error in classical conditioning. (A) Dopamine neuron activity in three circumstances was studied by Schultz et al. (1997). In the top example, dopaminergic neurons fire following an unexpected reward (R). In the middle example, the reward is predicted (CS), and the reward (R) occurs. In conformity with reinforcement learning theoretical predictions, the activity is unchanged. Finally, in the bottom case, neurons firing activity was reduced when a predicted reward predicted was omitted. (B) Three RL models components are considered for each scenario: the reward $r(t)$, the value-function $V(t)$ and the prediction-error $\delta(t)$. The yellow juice drop represents the reward (R), when the little red sun presents the conditioned stimulus (CS), i.e. when the reward is predicted following learning. In the top example, the reward occurs thus producing a positive prediction-error. In the middle example, after learning the association via the update of $V(t)$, the reward elicits a prediction-error of zero (the monkey is not surprised because the outcome is expected). In the bottom example, when the reward is omitted, the prediction-error becomes negative. Original figure is from Ludvig et al., 2011.

2.3.3 A neural common currency in humans

The relationship between changes in blood oxygenation of the brain and neural activity has been assumed since the end of the 19th century (Huettel et al., 2004). At the end of the 20th century, fMRI (Functional Magnetic Resonance Imaging) techniques were developing, and some studies started to investigate brain functioning by means of *BOLD* (Blood-Oxygen-Level Dependent) contrast in humans (Kwong et al., 1992).

In the decision-making field, most studies using fMRI involve a visual representation of a gambling

task, where subjects are instructed to choose between pairs of options by keypress. During the task, brain activity is monitored and thereafter correlated with various decision variables. fMRI thus permitted to study the neural coding of decision variables (prediction-error, reward, probabilities, etc.), and showed the prominent role of subcortical and cortical areas. Numerous studies highlight the role of the striatum and portions of pre-frontal cortex (PFC) and orbitofrontal cortex (OFC) in the coding of reward value (O'Doherty et al., 2004; Knutson et al., 2005; Daw and Doya, 2006; Tom et al., 2007a). Reward-prediction error signals for their part, are also frequently localized in the striatum, although some traces are found in OFC and amygdala (O'Doherty et al., 2004; Daw and Doya, 2006; Pessiglione et al., 2006; Yacubian et al., 2006). Distinguishing probability of reward from expected gain, was not an easy task, as most of the studies find entangled neural correlates for both variables (Delgado et al., 2005; Hsu et al., 2005; Knutson et al., 2005; Preusschoff et al., 2006).

Following these empirical evidences, the hypothesis that the brain does indeed make subjective expected value calculations was reinforced. This calculation is allegedly made via a set of sequential and modular processes (Fodor, 1983), within a two-step architecture. Roughly, each subprocess (e.g. valuation) is complete and isolated in a module and then passes on output information to the next module (e.g. choice) sequentially. The valuation stage was identified in the ventromedial prefrontal cortex and orbitofrontal cortex (vmPFC/OFC), parts of the striatum. The choice stage for its part, was found to be implemented in lateral PFC and other parietal areas (Rangel et al., 2008; Kable and Glimcher, 2009).

On this basis, the hypothesis of a *neural common currency* was formulated (Levy and Glimcher, 2012; Sescousse et al., 2013). More precisely, this hypothesis aimed at answering the following question: How can a decision-maker choose between options that are fundamentally different in nature? Utility theories, as well as RL models typically consider choices to be made as if the values of the options have been mapped to a single common scale. Conducting a meta-analysis using data from numerous fMRI studies, Levy & Glimcher (2012) proposed that this common representation

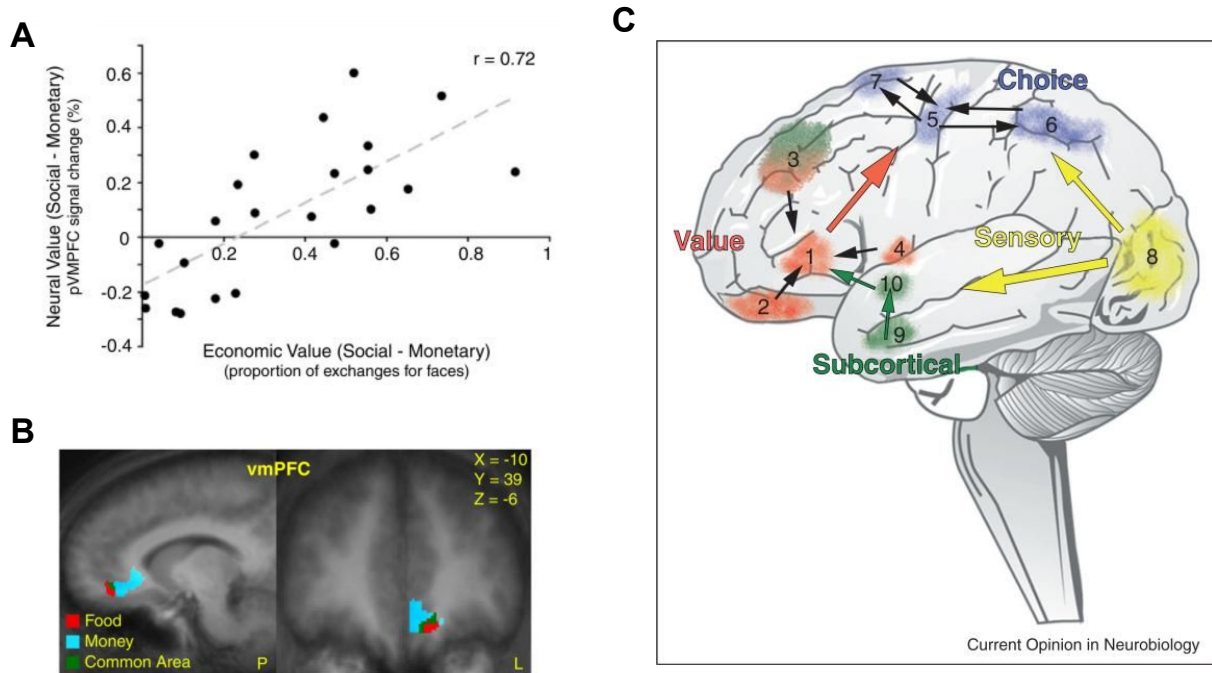


Figure 2.7: A valuation system for economic decisions. **(A)** A positive correlation between a neural value signal and behaviorally measured economic value was found in the posterior vmPFC. **(B)** A common area vmPFC/OFC was correlated with rewards of different natures, namely food and money. **(C)** A possible distributed circuit for value-based decisions. Information from occipital (8) and subcortical structures (9, 10) are used to build a single common value representation in prefrontal structures (1, 2, 3, 4). This unified value is then passed to motor cortical areas and parietal areas to produce the choice (5, 6, 7). (1) vmPFC, (2) OFC, (3) DLPFC, (4) Insula, (5) Primary motor cortex (M1), (6) Posterior parietal cortex, (7) frontal eye fields, (8) Visual cortex, (9) Amygdala, (10) Striatum. The original figures are from [Levy and Glimcher, 2012](#).

is located and dopaminergically driven in a subregion of the vmPFC and OFC (Fig. 2.7). Moreover, they show that different kind of supposedly incommensurable rewards (food and money) trigger vmPFC/OFC activations, that correlates with estimated subjective values. Other studies corroborated this overlap of primary and secondary rewards representations (e.g. [Delgado et al., 2011](#)).

Of note, this *valuation system hypothesis* is somewhat controversial. Indeed, value *per se* is often difficult to disentangle from other variables, as value often correlates with outcome identity, or higher-order cognitive phenomena such attention, arousal, salience, etc. ([Maunsell, 2004](#); [Heilbronner et al., 2011](#); [Schoenbaum et al., 2011](#); [Leathers and Olson, 2012](#); [O'Doherty, 2014](#)). In addition, the choice and valuation stage may not be anatomically dissociable, as shared areas are activated during both stages ([Bartra et al., 2013](#)). For these reasons, and among others, some schol-

ars have argued that behavioral research should foster other approaches to decision-making than value based (Suri et al., 2020; Hayden and Niv, 2021).

2.4 Summary

RL models are historically tied to animal learning and behaviorism (Thorndike, 1898; Skinner, 1938). Learning was conceived as the strengthening of association between stimuli (classical conditioning) or between stimuli and actions (operant conditioning) through reinforcers (a reward/a punishment). Behaviorism³, as psychology research program, advocated for a black box model of cognitive processes, where psychological entities should be disregarded, and reduced to inputs (stimuli), and outputs (behavior) (Box 2.1).

Later, computational models of classical conditioning emerged in an attempt to account for phenomena contradicting previous theories (e.g. blocking effect) observed in animal learning (Rescorla, 1972). Progressively, the association between a stimuli and other stimuli or actions, were formalized as a value function (Rescorla, 1972). A RL agent seeks to maximize its expected reward, by optimizing a state-value function (Sutton and Barto, 1981, 1987; Sutton, 1988; Sutton and Barto, 1990), or a state-action value function (Watkins, 1989). However some models actively avoid value functions in order to only focus on policy learning (e.g. policy gradient methods, see Bennett et al., 2021), claiming that value computation is not a parsimonious assumption (Hayden and Niv, 2021). In order to integrate new information and update the value function, an RL agent computes the prediction-error, i.e. the different between the obtained outcome and the expected outcome. The idea that the brain neurally implements such computation emerged when correlations between model variables and dopaminergic activity were found in monkeys (Schultz et al., 1997). Value based decision-making and RL thus started to be investigated in humans (Camerer et al., 2004b; Rangel et al., 2008; Rustichini, 2009) and the multi-armed bandit task became a standard paradigm

³Aside from particular behaviorists such as Tolman (1948)

to study it (Box. 2.3). In this paradigm, subjects have to rely on experience to learn the expected value of options (like in non-human animals), where contingencies (probabilities and outcomes) are not provided *a priori*.

Several scholars then turned proponents of the *neural common currency hypothesis* (Rangel et al., 2008; Levy and Glimcher, 2012; Sescousse et al., 2013). It posits that items of different natures can be compared through their mapping on a single common scale, and that this process is neurally represented. The brain is thought to represent the relevant decision variables, to compute value signals related to the variables at hand, and thereafter select the action possibility with the strongest value signal. This two-step model of value based decision-making has been tested through imagery methods. Indeed, various fMRI studies, identified a valuation circuit supposedly implemented in the orbitofrontal cortex, prefrontal cortex, striatum, and sometimes other cortical areas (e.g. Knutson et al., 2005; Pessiglione et al., 2006).

Interestingly, subjective value as conceived in psychology, has undergone an evolution similar to the concept of utility in economics. At first ontological speculations were avoided, but with the neuroscientific revolution the question of its physical implementation played an increasingly important role.

There are other reasons, too, for the incompleteness of logical contact that consistently characterizes paradigm debates. For example, since no paradigm ever solves all the problems it defines and since no two paradigms leave all the same problems unsolved, paradigm debates always involve the question: Which problems is it more significant to have solved?

Thomas Kuhn, *The Structure of Scientific Revolutions*, 1962

3

The description-experience gap

3.1 Evidence for a behavioral gap

3.1.1 Two lines of research

In the previous chapters we have seen two experimental research paradigms, that emerged concurrently, and led to different methods for studying decision-making as well as subjective valuation. In tasks involving description-based choices, individuals are presented with gambles, either described textually or graphically (e.g. pie-charts). Through the representation of the cue, individuals are directly provided with decision variables (probabilities, outcomes) *prior* to the actual

choice, i.e. decision under risk. In tasks involving experience-based choices, cues are abstract and are not supposed to convey any semantic meaning. They can take the form of symbols (e.g. a star) or day-to-day objects (e.g. a door). No information is given *a priori*, hence individuals must infer the expected value of each symbol by experience, i.e. by remembering past outcomes. In addition, they sometimes have to infer the underlying structure of the experiment, such as relation between states. Those decisions are considered under uncertainty (Knight, 1921) as outcome probability distributions are completely unknown. Although the paradigms are quite different in terms the models typically fitted (e.g. utility models against RL models) or theoretical assumptions regarding individuals' cognition – that reflects on experimental conditions – (e.g. quasi-full information against complete uncertainty), they share enough features to allow comparisons¹. Indeed, the multi-armed bandit metaphor (see Box 2.3) is rather close to the gambling metaphor of decision-making (see Box 1.4), and some modeling assumptions are common to both paradigms (e.g. value functions implied by the maximization of expected-value hypothesis). Thus, the description-experience dichotomy can be conceptualized as a continuum of uncertainty, rather than two binary categories (Hertwig and Erev, 2009b, Fig. 3.1).

3.1.2 First evidence

In the beginning of the 2000s, three studies pioneered the investigations of behavioral discrepancies between description- and experience-based choices (Barron and Erev, 2003b; Weber et al., 2004; Hertwig et al., 2004).

Barron and Erev (2003a) were among the first to re-parameterize description-based tasks to bring them closer to experience-based tasks². They did that by showing the outcome subsequent to the choice, and by repeating decision problems more than once. Their goal was to test the persistence

¹Luckily for us, these competing paradigms seemingly avoid being hit by the epistemological problem of incommensurability (Feyerabend et al., 1993; Kuhn, 2021).

²Erev (1998) had previously built bridges between the description and experience paradigm by fitting RL models to choices made in a game theory setting. Thaler, Tversky, Kahneman, and Schwartz (1997) and Fox and Tversky (1998) presented generalizations of prospect theory where agents have to rely on past experience. Even before Chu & Chu (1990) showed that preference reversals (Slovic and Lichtenstein, 1971) could be eliminated by presenting the outcome feedback subsequently to the choice

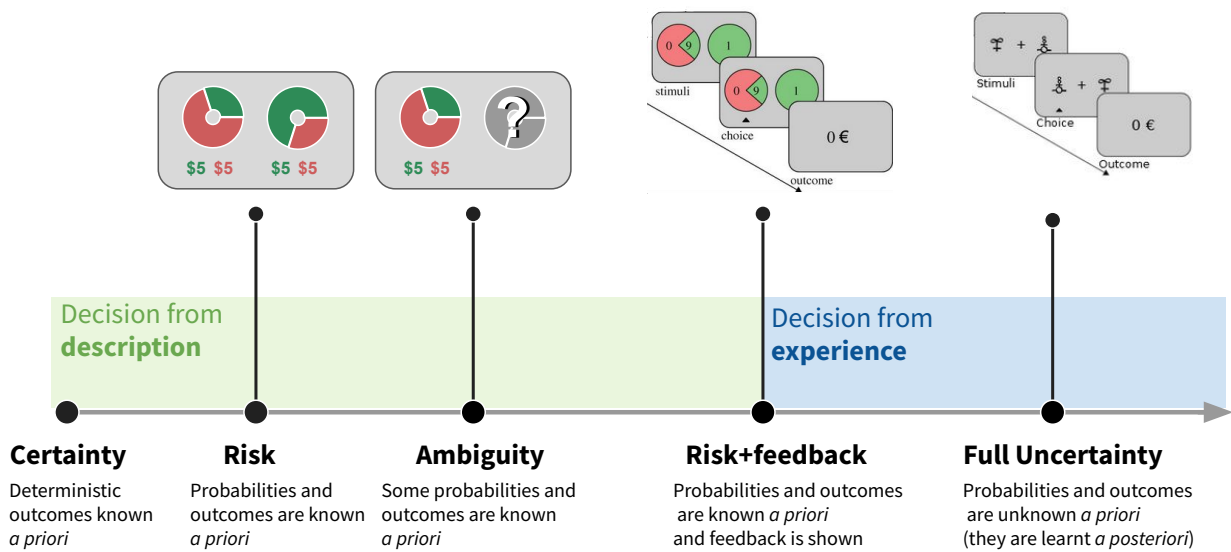


Figure 3.1: The continuum of uncertainty with regards to the description and experience paradigms.

of four common decision patterns predicted by prospect theory and observed in description-based tasks, namely, loss aversion, the certainty effect, the reflection effect, and the inverse s-shaped probability weighting function (Kahneman and Tversky, 1972). Aside from loss aversion, deviations from prospect theory were observed in all patterns.

First, the certainty effect was reversed. Initially, the certainty effect predicts that the riskier of two prospects is preferred if the probability of winning in both prospects is multiplied by a common ratio. This ‘common ratio’ effect constitutes a violation of expected utility theory’s (Allais, 1953) and was demonstrated by Kahneman and Tversky (1979) with the following gamble:

| Certainty effect |
|---|
| (L_1) 3 with certainty |
| (L_2) 4 with probability 0.8; 0 otherwise |

Most of the subjects here preferred the safe option (L_1). However, dividing each lottery by 4, which gives a probability of 0.25 for L_1 , and probability of 0.2 for L_2 , allowed to reverse the preference toward L_2 . Interestingly Barron and Erev noted that adding feedback led them to obtain an

opposite pattern (see also Jessup et al., 2008): subjects were risk-seeking in the first gamble (thus choosing L_2), whereas multiplying by a ratio (dividing) in the second gamble led to an increase of risk-averse choices (L_1).

Following the same procedure, i.e. presenting classical gambles from the literature but in an experiential learning setting, Barron and Erev showed that subjects exhibited two other departures from classically observed patterns in description-based choices. The reflection effect, that consists in being risk-seeking in losses and risk-averse in gains, was present but in an opposite configuration. Subjects tended to be more risk-seeking in gains than in losses. Similarly, the traditional overweighting of rare events observed in the inverse S-shaped probability weighting function gave way to an underweighting of rare events.

Weber et al. (2004), compared risk preferences across humans and animals. They noted that because animals are foreign to human symbolic and semantic representation of lotteries, all their decisions when foraging information are *de facto* decisions from experience. In order to mimic animal foraging tasks in humans, they used a *sampling paradigm* (Fig. 3.2). Essentially, it consists in 'foraging' the information before making a choice, by letting the subject sample each option (i.e. they decide how long to explore each option underlying outcome distribution) prior to the actual choice. They reported that when placed in an experiential sampling paradigm, humans and animals had similar risk attitudes, i.e. they tend to be more seeking in gains than prospect theory predicts.

Following Weber et al., Hertwig et al. (2004) investigated which properties of decision from experience caused the underweighting (instead of the traditional overweighting) of rare events. The two candidate properties were (i) the repeated decisions design (ii) the direct experience (or not) of the outcome subsequently to the choice. To disentangle those two properties, they mobilized the sampling paradigm from Weber et al. If underweighting of rare events is caused by repeated decisions, then decision by sampling should not display this bias. If it is caused by direct experience

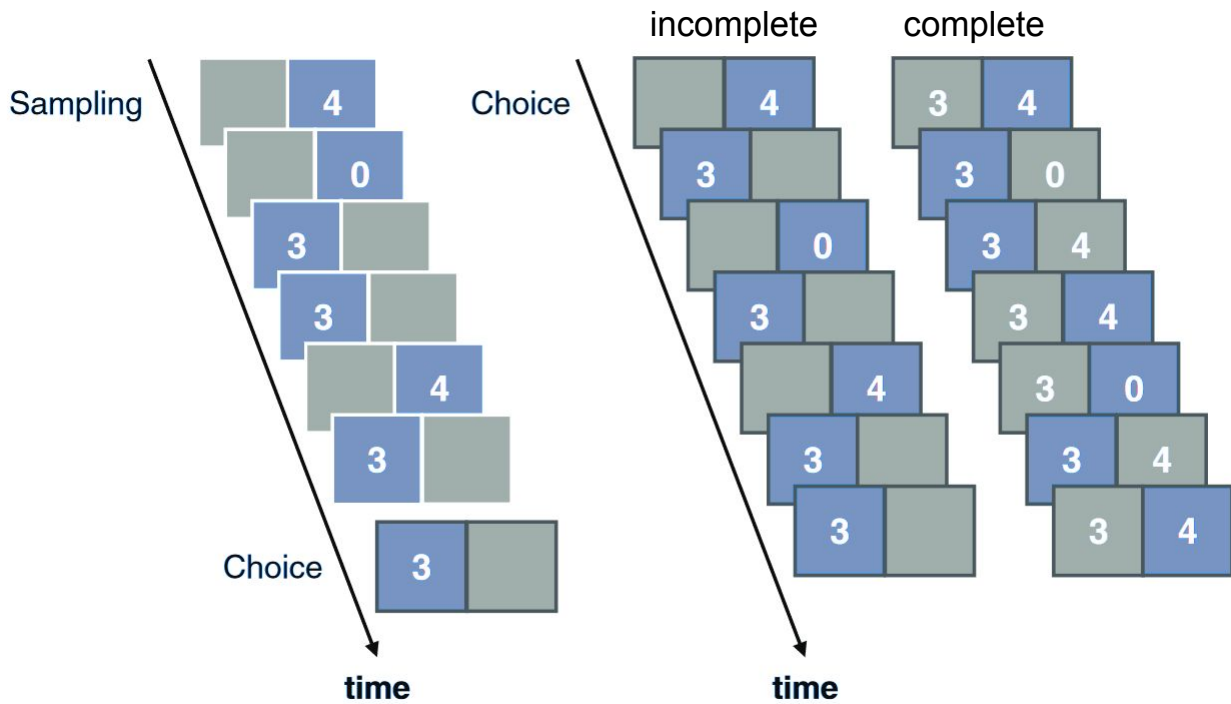


Figure 3.2: Different types of outcome presentation in experience. The sampling paradigm (leftmost) consists for the subject in a sampling phase, where he is able to explore options' possible outcomes before the actual choice. In the partial-feedback paradigm (middle), only the chosen option's outcome is shown subsequently to the choice. In the complete-feedback paradigm (rightmost), both options' outcomes are shown subsequently to choice. The figure is adapted from Hertwig and Erev, 2009b; Wulff et al., 2018.

of the outcome however, then the sampling paradigm should result in underweighting.

First they contrasted repeated decisions in experience and in description domain. They observed that in conformity with prospect theory predictions, subject overweight rare events in the description domain, while they are underweighted in experience. The 'repeated decisions' hypothesis being discarded, they conclude that the cause for such behavioral differences must lie in the outcome presentation. Finally, they identify two factors for the underweighting of rare events phenomenon: sampling error (i.e. subjects rely on small samples, such that the options seems less variable that they actually are) and recency (i.e. recently sampled outcomes are given greater weight than earlier sampled ones).

3.1.3 Testing the robustness of the description-experience gap

Strikingly, experience-based decisions seem to differ systematically from description-based decisions. In the experience domain, the fourfold pattern of risk attitudes has been found to hold, yet in opposite directions. This phenomenon where traditional patterns of description-based decision are reversed in an experiential learning setting is known as *the description-experience gap* (Hertwig and Erev, 2009a).

A meta-analysis conducted by Madan et al. (2014; 2019) initiated the study of the gap via cross-species comparisons. They compared risk-preferences among pigeons and humans, varying from described to experiential choices. In experience, both species presented an "extreme-outcome bias", i.e. both extremities of the probability distribution of experienced values were overweighted. Also, this extreme-outcome rule leads to a contextual and asymmetric treatment of gains and losses: relative gains elicit more risk-seeking attitudes. Again, experience-based choices showed an inverse *reflection effect*. Interestingly, when asked through a self-reported memory test, subjects recalled more occurrences of extreme outcomes than equally encountered non-extreme outcomes. Consequently, they suggest that this extreme-outcome rule (and thus partly the description-experience gap) is underpinned by specific cognitive processes, and in particular memory processes.

In order to test which model could account for this gap and its underlying decision processes. Model competitions were organized. The Technion competition (Erev et al., 2010) consisted in predicting risky choices within the description, partial feedback, complete feedback, and sampling paradigms (Fig. 3.2). A broad range of theoretical approaches were represented (e.g. heuristics, prospect theory, regression models, etc.). In the description paradigm, the winning model was a stochastic version of cumulative prospect theory (Tversky and Kahneman, 1992a). In the sampling paradigm and feedback paradigm however, winning models were respectively the ensemble

model and the ACT-R model. The former is a model combining several decision rules, such as the natural-mean heuristic (Hertwig and Pleskac, 2008) or the priority heuristic (Brandstätter et al., 2006). The latter is a model that implement a declarative memory system which can account for primacy and recency effects (Lovett et al., 1999). Consequently, models that capture behavior in description-based settings assume very different mechanisms than models capturing behavior in experience-based settings. This reinforces the idea that the *description-experience gap* is not only due to a difference in experimental paradigms but to different underlying cognitive processes.

Finally, Wulff et al. (2018) recently conducted an extensive meta-analysis in order to assess the robustness of the gap and identify its major determinants (Fig. 3.3).

They found that across the literature, identical decision problems presented in an experience- or description-based manners were leading to different preferences (Fig. 3.3A). They note that it is particularly true for decision problems that involve a risky against a safe option, where the gap is the most important.

The hypothesis stating that the gap (and in particular the underweighting of rare events) is partly caused by individuals relying on small samples (Barron and Erev, 2003b; Hertwig et al., 2004), is corroborated, as small samples distort experience-based representations of probabilities (Fig. 3.3B). Nevertheless, small sample reliance and thus sampling error, is not a sufficient explanation, as the gap persists even when experience and (objective) described frequencies converge.

In the sampling paradigm, the recency effect was replicated (Fig. 3.3C), yet for the effect to hold, the subject needs to have control over the information foraging process, i.e. it has to be its own decision to stop the exploration of possible outcomes.

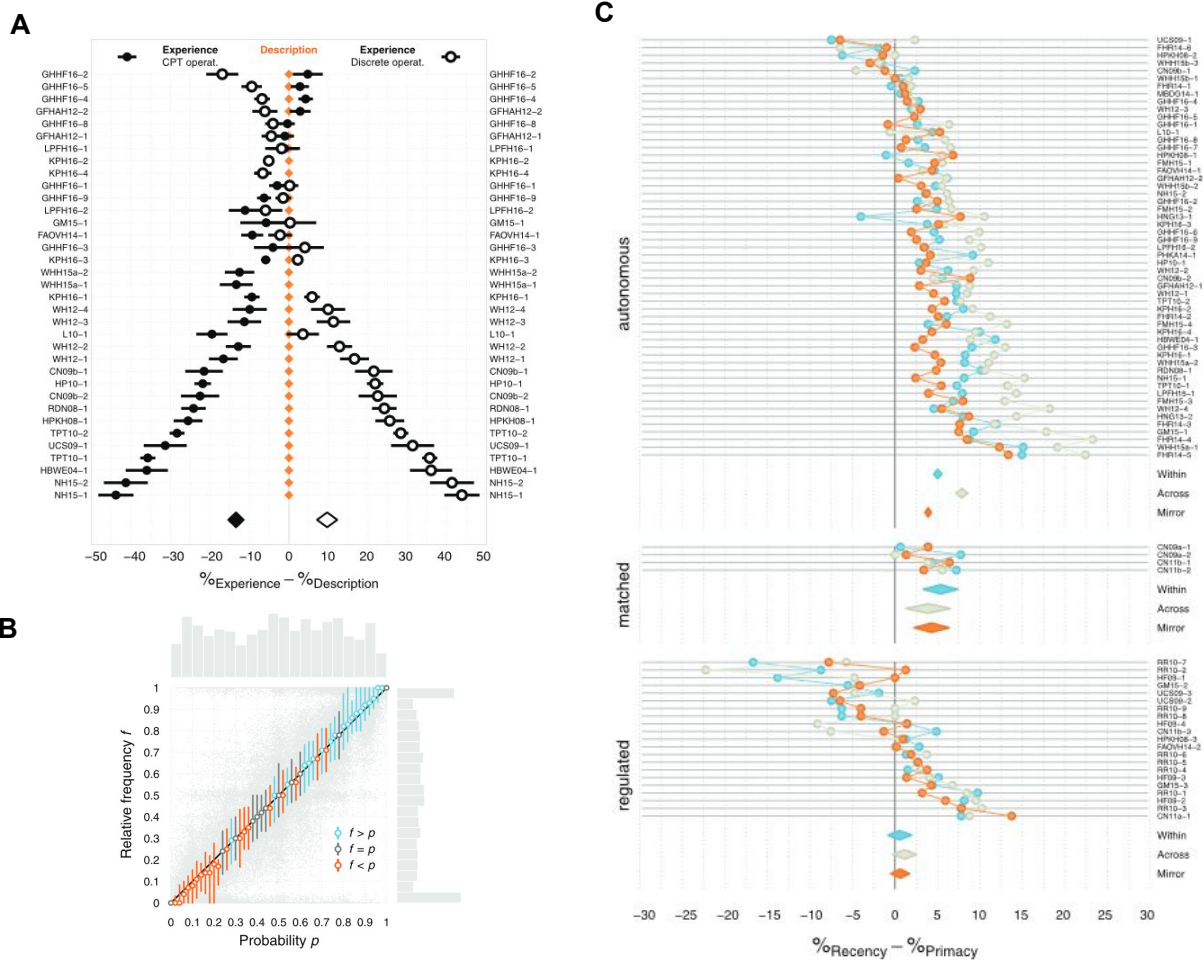


Figure 3.3: A meta-analysis of the description-experience gap. **(A)** The y-axis lists the 33 data sets that were analyzed. The x-axis quantifies the difference between the description and experience conditions. Each task had a description condition and an analogous experience condition. The gap was operationalized in two ways: either by quantifying the deviation from cumulative prospect theory predictions (left pane), either by merely considering discrepancy between choice proportions in the description and the experience condition (right pane). Error bars represent the standard error of the mean. **(B)** The x-axis represents the objective probability while the y-axis represents the experienced probabilities. Because of sampling error, experienced probabilities are distorted. Notably, the distribution of experienced relative frequencies is more polarized toward 0 and 1. The grey dots in the background are the individual trials. The circles and lines in the foreground (in blue, gray, or orange) represent the median experienced probabilities for each unique true probability and the respective interquartile range. The bar graphs at the top and on the right show the marginal distribution of the objective probabilities and of the experienced relative frequencies, respectively. **(C)** Analysis of the effect of recency in the sampling paradigm. In the autonomous (upper panel) and matched data sets (autonomous with pseudorandom sampling; middle panel), subjects sampled the options as much as they wanted. A strong recency effect was observed. In the regulated data sets (lower panel), subjects were forced to sample a certain N , and no recency effect occurred. Blue, beige, and orange points represent the results obtained for a given data set using three different measures (within-option, across-option, and mirror-image method). Diamonds and their widths represent the estimates and standard errors from a random effects meta-analysis. The figure is adapted from Wulff et al., 2018.

3.2 Objectives of the present work

3.2.1 First study

How should we investigate cognition, and consequently, decision-making? Cognitive phenomenon can be understood at different degrees of explanation (Marr and Poggio, 1976). Some academics claim that cognitive processes are best investigated and understood at a certain level, such as behavioral level, neural network level in the brain, etc. Others argue for a combination of top-down and bottom-up approaches, with emphasis on a particular level with regards to specific objects or questions (e.g. on interpreting neural data and brain functioning within behavioral paradigms, see Niv, 2021).

Non-invasive approaches for studying the human brain only allow macroscopic assessments of brain activity that aggregate thousands of cells (Glover, 2011). In contrast, animal models allow invasive recordings that give access to brain activity at the cellular and circuit levels. Taking advantage of the brain homology observed between humans and monkeys, decision under uncertainty and its neural mechanisms has often been studied in the latter (e.g. Hayden et al., 2011; De Petrillo et al., 2015). Prior studies showed that many biases replicate in monkeys, for instance ambiguity-aversion (Hayden et al., 2010; Rosati and Hare, 2011), or loss and framing effect (Chen et al., 2006; Krupenye et al., 2015). However, monkeys are by default unable to understand the symbolic language that allows humans to apprehend visual lotteries. As a result, they are compelled to learn a new symbolic system by reinforcement. Considering that the *description-experience gap* might be the consequence of different cognitive processes, ignoring this gap in monkeys could hinder our understanding of data coming from animal electrophysiology.

Thus, in a **first study** (chapter 4), we will assess to what extent the *description-experience gap* constitutes an epistemological challenge for decision-making research, in the sense that it could affect

the exchange of knowledge (top-down and bottom-up) between different levels of explanation, and therefore the building of plausible models of decision under uncertainty.

3.2.2 Second study

If the *description-experience gap* is more than a laboratory artefact, and originates from different cognitive processes or representations, does it have consequences for the *common currency hypothesis*? The *common currency hypothesis* (Rangel et al., 2008; Levy and Glimcher, 2012) posits that neural value representations are encoded in the brain as-if items' attributes are mapped into a single scale; which then allows comparison of options that are different in nature. However, as seen previously, studies of decision-making (from which common currency arose and is conceptualized in) often exclusively consider those two representational systems (experiential and symbolic) separately. Either values are external and conveyed through symbolic representations (pie-charts/text) or they are learnt through experience (abstract cues with no particular meaning or information). It is thus unclear whether experiential values that suppose internal representations can be further compared to values represented in the environment (Fig. 3.4).

To what extent does buying a lottery ticket (and associated probabilities of winning described in the back) involves a comparison of experiential against symbolic values? Does a choice between a food product we already experienced and a food product that displays an objective rating (e.g., nutriscore) involves a comparison between experiential and symbolic values? Are those two types of value even commensurable?

Prior studies have designed task with hybrid choices, in which descriptive and experiential information is combined (e.g., Erev et al., 2008; Erev et al., 2017; Lejarraga and Müller-Trede, 2017). However, few studies have included hybrid choices presenting experiential options (where expected value is learned by reinforcement) against symbolic options (which expected value is described and given prior to the choice).

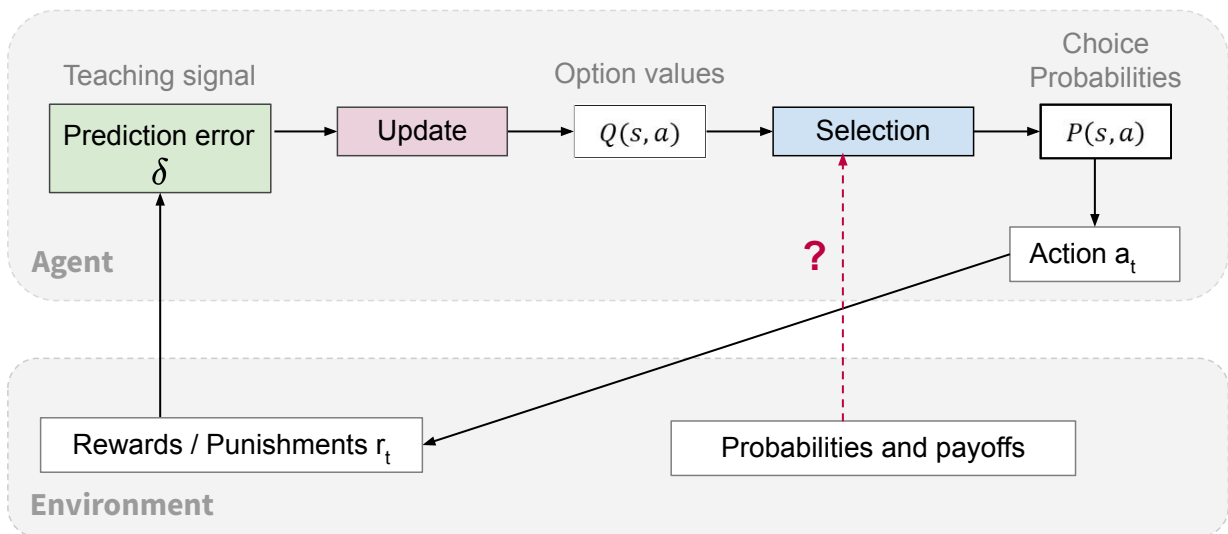


Figure 3.4: RL model with descriptions of objective probabilities and payoffs provided to the individual. Classically, an RL model selects actions in an environment. From the reward subsequent to the choice is computed a prediction error δ , which implements the difference between the obtained outcome and the expected outcome. It learns by updating its option-value function with the prediction error δ . Let's suppose an agent learned subjective values of uncertain options. Now, in a next phase, the previously learned options are presented against explicit options, which objective probabilities and payoffs *a priori*. Are the internal representations of uncertain options' values comparable to the explicitly described option values?

In monkeys, [Heilbronner and Hayden \(2016\)](#) presented experiential options against symbolic options and showed a preference for the former (at equal probabilities). In addition, they showed that as humans, monkeys are more risk-seeking in experience, suggesting that the gap also exists in monkeys. However, they only included five experienced options, which 3 are in the loss domain. This asymmetry does not allow an accurate assessment of the commensurability of the two types of value. In addition, it can be argued that the described options are of different nature to the ones in humans, as monkeys have to learn the symbolic system pertaining probabilities and outcome by experience.

In addition, [FitzGerald et al. \(2010\)](#) have looked for potential differential representations of experiential against symbolic values in humans. Activity in the vmPFC and OFC showed a positive response to learned value, replicating the previous literature for this specific valuation system.

On the other hand, neural substrates underpinning described values computation are not clearly highlighted, yet, the authors found activations in the bilateral ventral putamen and cerebellum. However, this study only includes three experiential option, against nine symbolic options. Also, because it is an fMRI study, the sample size is low. These two limitations prevent from assessing experiential option values precisely.

Thus, in a **second study** (chapter 5), we devised a series of experiments, which included almost 800 subjects in total. We aimed to accurately assess the degree of commensurability of experiential and symbolic values in a large population sample, while controlling for several biases.

4

The description–experience gap: a challenge
for the neuroeconomics of decision-making
under uncertainty

Review



Cite this article: Garcia B, Cerrotti F, Palminteri S. 2021 The description–experience gap: a challenge for the neuroeconomics of decision-making under uncertainty. *Phil. Trans. R. Soc. B* **376**: 20190665. <https://doi.org/10.1098/rstb.2019.0665>

Accepted: 30 July 2020

One contribution of 17 to a theme issue ‘Existence and prevalence of economic behaviours among non-human primates’.

Subject Areas:

neuroscience, behaviour, cognition

Keywords:

neuroeconomics, description–experience gap, reinforcement learning, decision-making, macaque, risk

Author for correspondence:

Stefano Palminteri

e-mail: stefano.palminteri@gmail.com

The description–experience gap: a challenge for the neuroeconomics of decision-making under uncertainty

Basile Garcia, Fabien Cerrotti and Stefano Palminteri

Laboratoire de Neurosciences Cognitives et Computationnelles, Ecole Normale Supérieure, Institut National de la Santé et Recherche Médicale, Université de Recherche Paris Sciences et Lettres, Paris, France

SP, 0000-0001-5768-6646

The experimental investigation of decision-making in humans relies on two distinct types of paradigms, involving either description- or experience-based choices. In description-based paradigms, decision variables (i.e. payoffs and probabilities) are explicitly communicated by means of symbols. In experience-based paradigms decision variables are learnt from trial-by-trial feedback. In the decision-making literature, ‘description–experience gap’ refers to the fact that different biases are observed in the two experimental paradigms. Remarkably, well-documented biases of description-based choices, such as under-weighting of rare events and loss aversion, do not apply to experience-based decisions. Here, we argue that the description–experience gap represents a major challenge, not only to current decision theories, but also to the neuroeconomics research framework, which relies heavily on the translation of neurophysiological findings between human and non-human primate research. In fact, most non-human primate neurophysiological research relies on behavioural designs that share features of both description- and experience-based choices. As a consequence, it is unclear whether the neural mechanisms built from non-human primate electrophysiology should be linked to description-based or experience-based decision-making processes. The picture is further complicated by additional methodological gaps between human and non-human primate neuroscience research. After analysing these methodological challenges, we conclude proposing new lines of research to address them.

This article is part of the theme issue ‘Existence and prevalence of economic behaviours among non-human primates’.

1. The neuroeconomic research programme

The expected utility model was established as the standard normative model of decision-making under risk [1,2]. Integrating Bernoulli’s intuition about the curvature of the utility function and probability theories, von Neumann and Morgenstern demonstrated that choices based on the expected utility (i.e. the product between the utility of an outcome and its probability) satisfies four basic axioms of rationality (completeness, transitivity, continuity and independence). Historically, the neoclassical economics research programme disregarded the study of the internal processes governing economic behaviours. Keynes’ animal spirits [3] were considered unmeasurable, and economic theory was built on the assumption that the human mind as well the brain were ultimately black boxes. The ‘as-if’ hypothesis [4] illustrates this position by endorsing an instrumentalist epistemology: theory predictive power prevails on the realism of its initial assumptions. Accordingly, it was considered acceptable to rely on unrealistic assumptions regarding the unbounded cognitive capacities or perfect knowledge of economic agents, as far as the predictions were sufficiently accurate.

However, with the accumulation of behavioural evidence against the standard normative expected utility model, it soon appeared that it had to be profoundly amended to successfully account for actual decisions under risk [5,6]. Positive, descriptive, models of decision-making under risk that integrate insights from psychology, such as the notion of bounded rationality (i.e. humans display limited computational capacities), heuristics (taking computational *shortcuts* to make decisions) and biases (systematically distorted representations of behavioural variables) were then proposed and formalized [7–9]. Among the descriptive theories of decision under risk and uncertainty, ‘prospect theory’ (PT) had a strong empirical ground and stood out [8,10]. PT postulates that expected utility is calculated relative to a reference point (the *frame*), an asymmetric treatment of gains and losses (*loss aversion*), as well as a subjective weighting of probabilities (*probability distortion*). PT successfully explained known paradoxes (such as the Allais’s paradoxes) and new ones (e.g. the Asian disease paradox, as well as a certain number of ‘real life’ irrational behaviours [11,12]).

However, despite these successes, some aspects of the descriptive approach, in general, and PT, in particular, remained unsatisfactory. First, it remained difficult to ultimately arbitrate between competing descriptive theories solely based on behavioural data. For instance, alternative behavioural theories have been proposed (such as rank-dependent utility, regret and disappointment theories; see [13] for a review) that make overlapping predictions with PT, making them hard to disentangle. Second, while making accurate predictions, PT, and other descriptive theories, do not specify which are the actual cognitive operations and how they are implemented by the brain. In terms of the Marrian analysis of modelling, PT (as other descriptive theories) is situated at the *computational* level that specifies which is the goal of the agent (in this case: maximizing a subjective utility that includes reference point dependence, loss aversion and probability deformation), but is silent concerning the *algorithmic* (i.e. what are the operations involved in the manipulation of decision variables) and *implementational* levels (i.e. how these operations are physically embodied and realized) [14].

A couple of decades later the time was ripe for a group of scholars of diverse origins to seek in neuroscientific data the way to overcome the limitations of descriptive theories, developed by psychologists and behavioural economists. This was facilitated by the rapid development of non-invasive neuroimaging techniques in humans (most notably functional magnetic resonance imaging: fMRI [15–17]) and improvement of single-unit electrophysiological recordings in monkeys [18,19]. The hope was (and still is) that, taking advantage of neuroscientific methods and concepts, *neuroeconomics* (as this raising field was named), would be able to address the epistemological issues of economic theories highlighted above.

Concerning adjudicating on competing theories (our first issue), by opening the brain ‘black box’ functional neuroimaging studies would provide an additional crucial observable measure—blood oxygen level dependent signal (BOLD: an aggregate and indirect measure of neural electrical activity), to compare, falsify and ultimately refine behavioural models. We define this approach as the *weak neuroeconomic agenda*, as it does not involve rewriting economic descriptive theories [20–22]. Coming back to our example, while making similar behavioural predictions in respect of preferences under risk,

different theories postulate different utility functions that can be searched in the brain [23–25]. Assuming one knows where to look for utility representation in the brain,¹ it would be, in principle, possible to assess which model better predicts its activity (a sort of neural model comparison: see [29]). Beyond comparing different theories, the neural activity could in principle help refining a theory by fixing some of its parameters. For instance, in many circumstances, PT is silent about how the reference point should be set [30]. Assuming one knows where to look for positive (gain) and negative (loss) utility representations in the brain, in some cases the reference point could be inferred comparing the profile of activity of the ‘gains’ and ‘losses’ areas² [25,33].

Concerning building new theories (second issue), accepting the fundamental ontological tenet that (economic) decisions ultimately stem from neural activity in the brain (which is a standard materialistic and monistic solution to the mind-body problem, see [34]), entails that neuroscientific methods should provide the conceptual and methodological tools necessary to develop new, neurobiologically grounded, *neural* models encompassing the algorithmic and implementational levels. By contrast with the previous approach, we define this approach as the *strong neuroeconomic agenda*, as it involves rewriting economic theories in neurobiological terms. By integrating biological constraints and cost functions, these hypothetical neurobiologically grounded economic models have the potential of explaining why human decision-making presents certain biases from a biologically (not logically or statistically) normative perspective [35,36].

The methodological requirements of the two main neuroeconomics agenda are not quite the same. The weak neuroeconomic agenda can, in principle, be fulfilled by experiments relying on aggregate and indirect measures of the neural activity, such as the BOLD signal recorded by fMRI scanners in areas encoding subjective values. Furthermore, since the goal is arbitrating between different behavioural theories of decision-making developed by psychologists and economists, the experiments belonging to this research agenda should be preferentially (if not exclusively) performed in humans.

On the other side, as neural models are, ultimately, models of which information is encoded in neurons and how neurons are connected (networks), the strong neuroeconomic agenda research programme cannot be pursued only relying on fMRI neural signals.³ In fact, BOLD signal, at its best resolution, aggregates over thousands of neurons [37–39]. Furthermore, it is still unclear to which extent it reflects presynaptic or postsynaptic activity (probably a mixture of both) [39,40]. Such neural models should eventually be validated based on the recording of single-cell activities, which is, for obvious ethical reasons, nearly impossible in humans.⁴ This is why neuroeconomics research, from its very inception, strongly relies on electrophysiological research on animal models, which have been employed in the study of neural mechanisms and cognition for almost 80 years [42]. Monkeys (especially rhesus monkey: *Macaca mulatta*), are particularly popular models, because they present a wide behavioural repertoire and high degree of neuro-anatomical homology with humans, especially concerning the prefrontal cortices that underpin decision-making [43].

In figure 1, we represent what a prototypical workflow should look like to combine human and monkey data to deliver a neural model of decision-making. Of note, we describe it from an abstract perspective of theory-building, but in

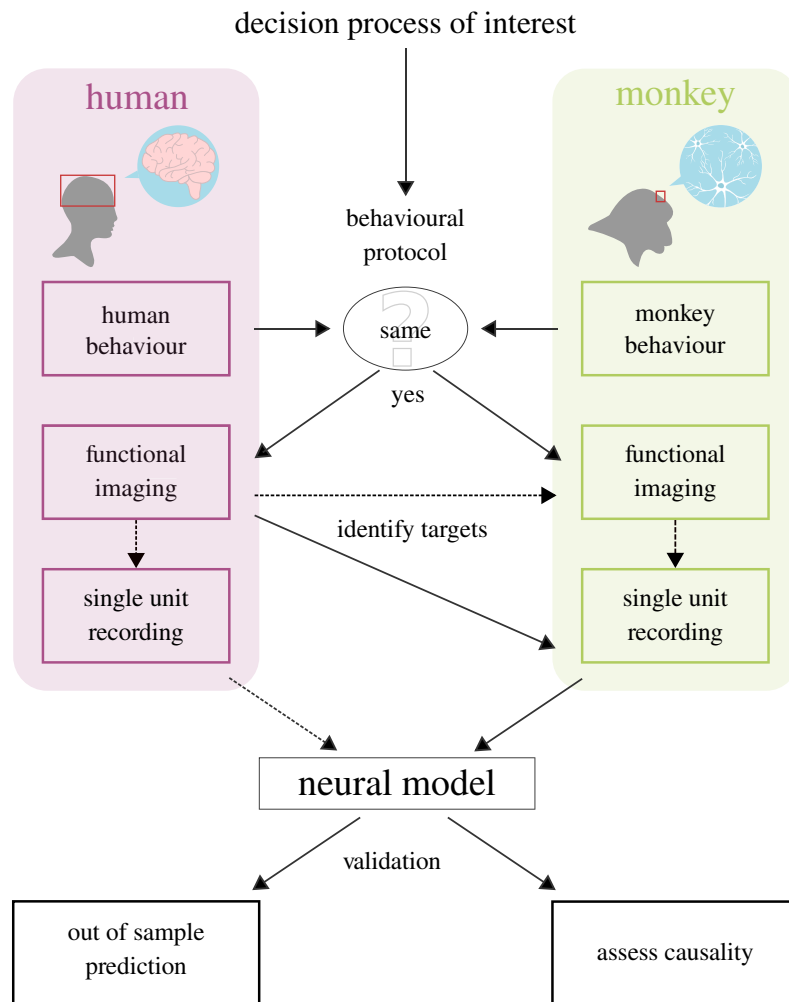


Figure 1. Prototypical workflow combining human (purple) and monkey (green) data to pursue the strong neuroeconomic agenda. Dotted lines designate optional steps. (Online version in colour.)

reality, its different steps can occur simultaneously (or in reverse order), and in very distant laboratories. Once identified as a behavioural process of interest (e.g. decision-making under uncertainty), a behavioural protocol is designed (typically, a series of choice problems involving different amounts of rewards and probabilities) and administered to both humans and monkeys. If the behaviour is comparable across species (meaning that the monkey represents a valid *experimental model* of human behaviour⁵), functional imaging in humans can then be deployed to identify neural targets encoding macroscopic variables (e.g. probabilities, outcomes) that are later used to guide the selection of the areas where neurons will be recorded in monkeys. A desirable intermediate step, to reinforce the functional correspondence between human and monkey brain activations, would be to also deploy fMRI in monkeys [45]. Similarly, in some neurologic and psychiatric diseases, intra-cranial neural activity can also be recorded in humans [41]. Finally, all these data can then be combined together to propose and validate a neurobiologically plausible model of the behavioural process of interest. Thereafter, the proposed model should be validated using lesions and assessing its generalizability. Methods such as trans-cranial magnetic stimulation and brain lesions can be used to test the alleged causal relationship between neural correlates and behavioural processes [46–48]. The model's ability to generalize can be assessed by generating predictions in tasks involving different decision problems and behavioural processes (out-of-sample validation).

A crucial step in this workflow is checking that humans and monkeys display the *same* behavioural processes and biases as a result of a true homology. This is something notoriously tricky to assess, because several, to some extent unavoidable, methodological differences exist between human and non-human primate research.

The foundational experimental paradigm of behavioural decision-making research consists in making choices between 'lotteries' or 'gamblers', i.e. options associated with known or unknown probabilities of obtaining different outcomes [2,5]. According to the gambling metaphor of individual choice [49], lotteries are believed to be prototypical of real-life decisions [50]. Outcomes and their probabilities are described to participants, who often (especially in the first generation of behavioural economics studies) make only one or very few decisions, without being informed about the outcome of their choices (in general to purposely prevent learning processes from influencing decision-making [51]). On the other side, monkey electrophysiological research adopts very different methodological standards. For various reasons (including ethical ones), monkey studies are limited in terms of sample size, and consequently, the number of observations per subject is greatly increased in order to increase statistical power and reduce measurement noise. In fact, behavioural tasks in monkeys display a greater number of trials per subject, collected on a sample size of often less than five subjects (e.g. [52,53]). Both parameters (sample size and number of trials) are roughly a couple of orders of magnitude different compared

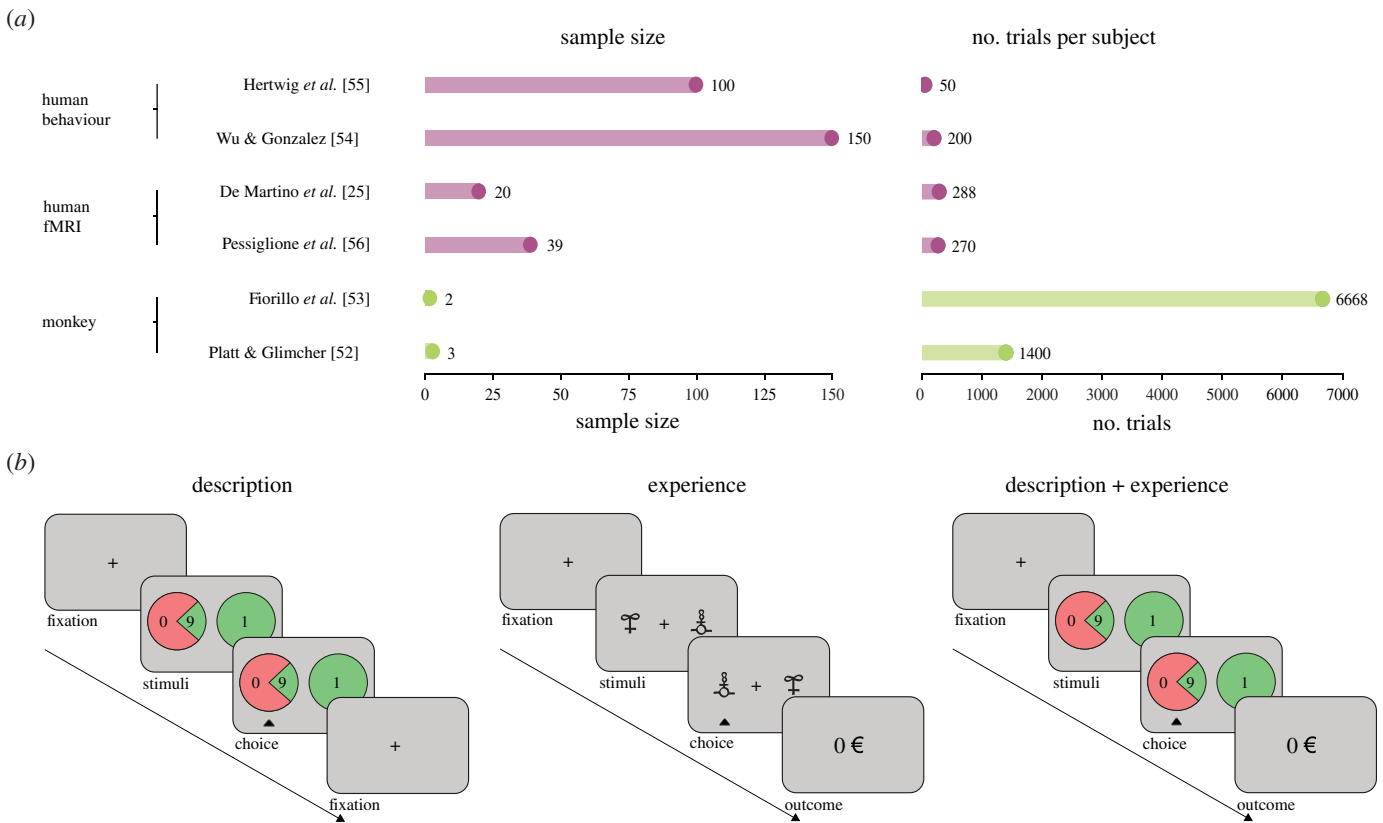


Figure 2. Methodological differences between description, experience and description–experience studies. (a) Sample size and number of trials listed in two electrophysiological studies [52,53], two human fMRI studies [25,56] and two human behavioural studies [54,55]. (b) Successive screens of a trial in the different behavioural decision-making paradigms. In pure ‘description’ paradigms, decision variables are explicitly described and no feedback is provided. In pure ‘experience’ paradigms, decision variables are hidden and feedback is provided on a trial-by-trial basis. In the ‘description plus experience’ paradigms, decision variables are explicitly described and feedback is provided on a trial-by-trial basis. (Online version in colour.)

to what is common practice in behavioural economics (e.g. [54,55]) (figure 2a). Interestingly, fMRI studies of decision-making present experimental parameters somehow in-between those used in monkeys and human studies: they usually involve hundreds of trials and also sample sizes of about 20–40 subjects (see two notable examples in neuroeconomics: [25,56]). Assuming that decision-making possesses ergodicity (i.e. the behaviour averaged across trials is the same as the behaviour averaged across subjects), different ratio trial/participants *per se* should not present a big challenge to compare results from human and monkey studies (but note that ergodicity does not seem to be granted for psychological processes, see [57]). However, in addition to these quantitative differences, in monkey studies, an outcome (usually a primary *reward*) is provided on a trial-by-trial basis. This is because a monkey would simply stop doing the experiment in the absence of extrinsic motivation. Thus, in virtually all cases monkey experiments include a *reinforcement learning* component, where actions are associated with past outcomes. This is true even when the paradigm involves establishing a symbolic system to communicate outcomes and probabilities. In fact, in the absence of a shared language or semantic system to communicate, monkeys are compelled to learn any representational system by trial-and-error from feedback.

In the present article, we argue that the above-mentioned differences do not only present a *technical issue*, but also a major *epistemological challenge* for the (strong) neuroeconomic agenda. We detail why below.

2. The experience–description gap

As mentioned before, foundational contributions to behavioural decision-making research were made through the use of explicitly described gambles. Several representations have been used to convey outcome values and probabilities, including textual and numerical descriptions (e.g. [5,8,54]), later replaced by visual cues such as pie-charts (e.g. [25,58]). In these paradigms, the information pertaining to the decision-relevant variables is processed by verbal and mental calculation systems and relies upon some degree of semantic knowledge to decode the meaning of the symbols used. In addition to that, decision problems were usually presented only once and, in case multiple decision problems were used, the final outcome (i.e. the realization of the lottery) was usually not displayed on a trial-by-trial basis (figure 2b).

However, relatively few situations in real life match the characteristics of the pure *description-based* paradigms, namely complete and explicit information about outcome values and probabilities. In fact, in many circumstances, it seems rather prudent to assume that information about outcome values and probabilities are shaped by past encounters of the same decision problem. Experimentally, this configuration is often translated into *multi-armed bandit* problems (starting with Thompson [59], but see [60] for a review), where the decision-maker faces abstract cues of unknown value and has to figure by trial-and-error the value of the options. Computationally, behaviour in multi-armed bandit problems is generally well-captured by associative or reinforcement learning processes

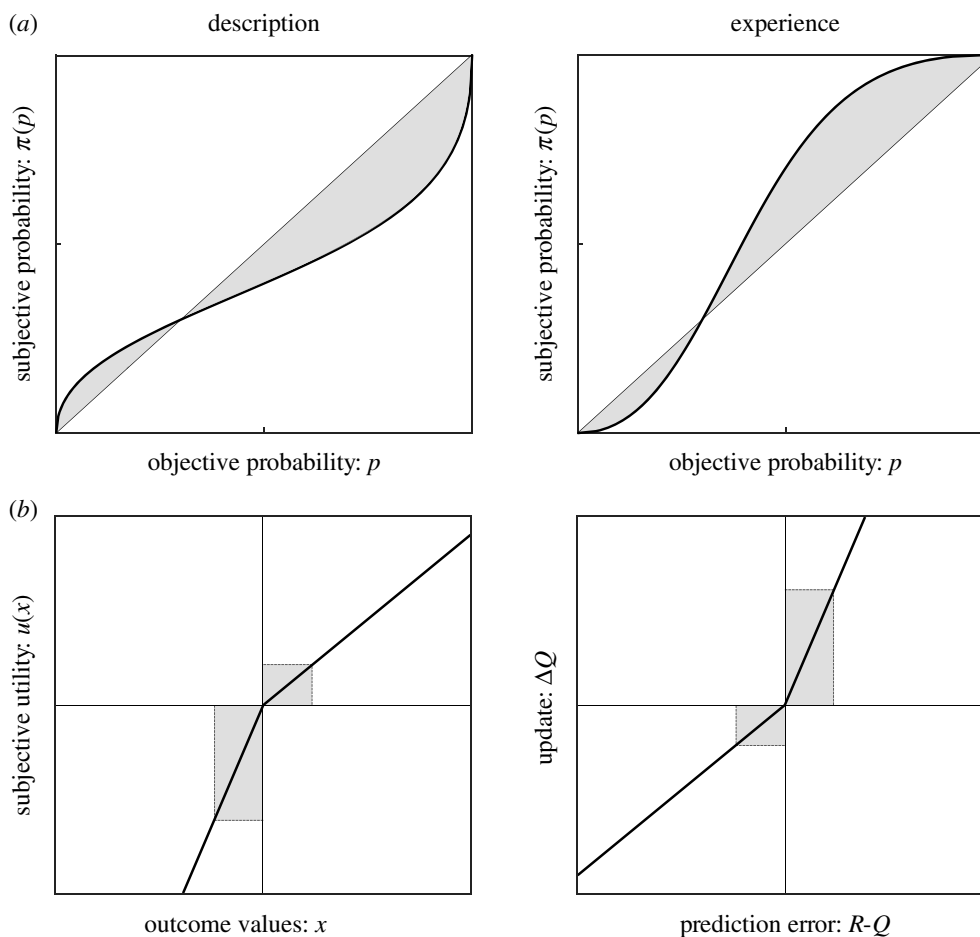


Figure 3. (a) Illustration of the nonlinear transformation of probabilities in description (left panel) and experience (right panel). In the description domain, subjective probability is reflected by a probability weighting function (here denoted π) following an inverse S-shape (i.e. low probabilities are overweighted while high probabilities are underweighted). This tendency is reversed when it comes to the experience domain, where the curve follows an S-shape. (b) Illustration of the classical linear utility function in the description domain (left panel) and the update of the value function for the experience domain (right panel). In description, the utility curve displays a steeper slope for losses than gains. In experience, an opposite phenomenon is frequently observed. The sign of the prediction error (i.e. the difference between the obtained reward R and the associative value Q) affects the learning rate.

[61]. In the early 2000s, a line of enquiry arose where researcher translated the typical decision problems used in behavioural economics (i.e. involving choices between a safe and a risky prospect in the gain and loss domain⁶) into experience-based paradigms [55,63,64] (figure 2b). Systematic comparisons between these two decision-making modes revealed the existence of robust *description–experience gaps* regarding risk preferences in humans [65–67]. More precisely, probability weighting functions eventually show opposite deformations when comparing description-based and experience-based choices (figure 3, box 1). In particular, most of the tenets of PT do not seem to hold in experience-based choices [8]. While traditionally, in the description domain, the occurrence of rare events is overestimated (*possibility effect*) and the occurrence of frequent events is underestimated, experience-based decisions tend to show the opposite biases: an effect that is only partially explained by incomplete sampling [55,63,64,66].

In description-based choices, a behavioural hallmark of loss aversion (overweighting of negative outcomes) is the *reflection effect*, where subjects are risk averse in the gain domain and risk seeking in the loss domain. The opposite pattern has been repeatedly found in the experience-based decisions [67]. This observation may be explained by biases in the learning process, such as remembering preferentially extreme outcomes or integrating preferentially better-than-expected outcomes [72,77]. Finally, a smaller subset of studies investigated a hybrid situation where decision problems are fully described, choices are repeated and followed by a trial-by-trial feedback. These

‘description plus experience’ paradigms showed that probability distortions compatible with prospect theory are initially present, but corrected by the presence of feedback [78,79]. To summarize, the whole spectrum of decision-making under uncertainty in humans is far from being fully captured by PT’s loss aversion and subjective probability deformation. Specifically, different descriptive models seem to apply as a function of how outcome and probability information is conveyed. In what remains of the paper, we illustrate why we believe that this feature seriously challenges leveraging on neural and behavioural data in monkeys to build a neural model of decision-making under uncertainty.

3. Decision under risk in monkeys

In this section, we try to address the question of whether monkeys are a good experimental model for human decision-making under uncertainty. We will focus this survey on rhesus monkey (*Macaca mulatta*) results because most electrophysiological studies are performed in this species (but see [44] for a more detailed review including other primates). Asking whether monkeys are a good experimental model translates into asking whether in the laboratory setting their behaviour displays the distinctive features and biases observed in humans. We stress again that the comparison is complicated by the fact that *pure* description-based paradigms cannot exist in monkey studies because of the lack of language. In fact, in

Box 1. Description- and experience-based behavioural models.

In this box, we sketch the formalisms standardly employed to explain and quantify risk preferences in description-based and experience-based decisions. Description and experience paradigms radically differ in how they model decision under risk. In the description domain, risk preferences are the *direct* result of subjective deformations of probabilities and outcomes that are explicitly stated. On the other side, in the experience domain there is no separate representation of outcomes' probabilities and no explicit deformation of outcomes' values. Consequently, risk preferences are the *indirect* result of the learning process that links past outcome information to subsequent choices. Eventually, these two approaches lead to different explanations of risk attitudes.

Risk preferences in description-based paradigms are commonly explained by prospect theory (PT). The expected value of k iterations of the same gamble X (which is random variable) is computed as follows:

$$E(X) = \sum_{i=1}^k p_i x_i,$$

where x_i is the value of an individual outcome and p_i is the objective probability of the outcome. PT states that the utility of an outcome, that is the subjective value $u(x_i)$, is nonlinear and modulated by different parameters: α and β , that are the power to which, respectively, a positive or negative outcome are elevated, and λ the loss aversion coefficient. Thus, the PT utility function is defined as follows:

$$u(x_i) = \begin{cases} x_i^\alpha & \text{if } x_i \geq 0 \\ -\lambda(-x_i)^\beta & \text{if } x_i < 0 \end{cases}$$

an $\alpha \leq 1$ corresponds to risk aversion in the gain domain (the intuition dates back to Bernoulli), $\alpha > 1$ corresponds to risk-seeking behaviours. In the loss domain, the same relation is true concerning the values of β . A value of $\lambda > 1$ corresponds to loss aversion; its typical empirical value is around 2 [10,68]. A decision-maker with $\alpha < 1$, $\beta > 1$ and $\lambda > 1$ will present different risk preference in the gain (risk aversion) and the loss (risk seeking) domain (figure 3b).

In addition, PT postulates a subjective deformation of probabilities. There are multiple ways to mathematically express the probability weighting function. One of the most common is the 'Prelec' function [69]:

$$\pi(p_i) = e^{-\delta(-\log(p_i))^\gamma}$$

with δ controlling the elevation and γ the curvature. When both parameters are set to 1, the function tends to linearity. The more $\gamma > 1$, the more the function adopts an S-shape. A classical result is the overweighting of low probabilities compared to high probabilities, where the direction of the curve follows an inverse S-shape (figure 3a), with $\gamma < 1$. Note that another probability weighting function has been proposed [54]. Finally, the subjective expected utility is given by

$$SEU(X) = \sum_{i=1}^k \pi(p_i) u(x_i).$$

By the variation of these parameters, PT accounts for inter-individual differences in risk preferences. Of note, concurrent theories such as regret theory [70] or rank-dependent utility models [71], which use very different representational structures and parameterizations, are also used to model decision-making under risk.

Experience-based paradigms can be seen as reinforcement learning problems operationalized as k-armed bandit tasks [61]. Consider an environment composed by a state vector S , with $s \in S$. In each of states s , there are available actions denoted $a \in A$. Each state-action pair has an underlying reward probability distribution, such that $P[R|s, a]$, is the probability of obtaining the reward R , knowing the state-action couple (s, a) . An agent must then follow a policy in order to maximize a state-action value function $Q(s, a)$ (i.e. to maximize the average expected reward). A common learning policy is to compute subsequently to a choice of the prediction error δ , that will be used to incrementally update the value associated to a specific state-action pair (s, a) :

$$\begin{aligned} \delta &= R - Q(s, a) \\ Q(s, a) &\leftarrow Q(s, a) + \alpha \delta \end{aligned}$$

with α the learning rate that determines to what extent newly acquired information overrides the previous. In this paradigm, inter-individual variability in behaviours can be accounted for by differences in individual parameters such as the aforementioned learning rate α . However, this model with only one parameter is too simple to accommodate different risk preferences.

A way to refine this model to account for different risk preferences, is to allow for two different learning rates, α^+ and α^- :

$$Q(s, a) \leftarrow Q(s, a) + \begin{cases} \alpha^+ \delta & \text{if } \delta > 0 \\ \alpha^- \delta & \text{if } \delta < 0 \end{cases}$$

If $\alpha^+ = \alpha^-$, the two learning rates model is equivalent to a one learning rate model. We define the tendency to preferentially update $Q(s, a)$ from positive prediction errors rather than negative prediction errors as *positivity bias* (or *loss neglect*) ($\alpha^+ > \alpha^-$). Conversely, we define the opposite situation ($\alpha^+ < \alpha^-$) as *negativity bias* (or *loss enhancement*).

The learning rate asymmetry has direct consequence for risk preferences in the setting where a subject has to learn the value of a safe (say a fixed value of 0) and a risky (say 50% chance of winning/losing one euro) option. A subject displaying a positivity bias would neglect the past losses and will, therefore, be a risk-seeker (figure 3b). Conversely, the negativity bias

implies risk aversion. Both pessimistic and optimistic biases have been reported in the literature, with the latter bias being more frequently reported [72–75].

While it is tempting to see the positivity bias as the experience-based antithesis of loss aversion, their formalism and psychological interpretations are quite different and they are, therefore, not mutually exclusive. Indeed, loss aversion concerns the valuation of *prospective* losses, while the positivity bias concerns the *retrospective* assessment of past losses.

It is important to note that, in humans, although the average values of the behavioural biases are reported as described above (for instance: inverse S-shape in description-based paradigms and loss neglect in experience-based paradigms; see figure 3a), their results are further tempered by a high degree of inter-individual variability in the bias parameters. At the individual level, some subjects may in fact display opposite biases in both experimental settings [72,76]. If inter-individual variability is equally high in other primates, the fact that monkey studies use very small sample size (figure 2) can contribute to explaining the comparably less consistent picture observed (table 1).

monkey studies, whenever outcomes and probabilities are conveyed via a symbolic system, the system is nonetheless learned and maintained by trial-by-trial outcomes (i.e. a situation similar to the ‘*description plus experience*’ paradigm, described above). In such ‘pseudo’ description-based paradigm, monkeys are trained to associate continuous variations in one visual feature (e.g. colour or size) to continuous variations of a decision variable (e.g. outcomes or probabilities). The comparison is further complicated by the fact that only few studies formalize risk preferences in terms of model parameters (such as probability distortion, loss aversion or learning rates) and data reporting is often limited to behavioural measures.

The general picture (table 1) emerging from ‘pseudo’ description-based paradigms in monkeys (i.e. studies relying on learned symbolic systems to communicate values) is, at best, mixed. PT has been explicitly tested in paradigms using visual cues carrying symbolic information similar to those presented to humans (e.g. pie-charts). Only a few studies show results in conformity with the pattern of description-based decisions observed in humans. Risk aversion, suggestive of marginally decreasing utility in the gain domain, has been rarely reported [93]. Nioche *et al.* [98] is the sole study confirming all PT features: marginally decreasing utility (risk aversion in the gain domain), loss aversion (risk seeking in the loss domain) and subjective probability weighting consistent with overestimation of rare events. Probability weighting function consistent with standard PT has been reported by other studies, but the same studies also reported increasing marginal utility and risk seeking in the gain domain, which is not typically observed in description-based decisions in humans [95,97]. Many others pseudo description-based experiments also reported risk-seeking attitudes and/or marginally increasing utility in gains [91,92,94,96]. In addition, although the traditional inverse probability weighting function has sometimes been observed [95,98], variation of experimental design features (such as randomly mixing gambles instead of repeating the same gambles sequentially) can reverse the direction of the probability weighting function [99].

Regarding ‘pure’ experience-based studies in monkeys (i.e. involving no symbolic system to communicate values), the picture is somehow clearer. Indeed, rhesus macaques exhibit robust risk-seeking behaviour in the gain domain [80–89]. Risk-seeking attitudes have also been reported in the loss domain [90].

Risk-seeking behaviour in experience-based studies can be computationally explained by an increased sensitivity to positive (compared to negative) prediction errors (‘positivity’ bias) which is generally documented in human reinforcement learning (box 1) [72–74]. This hypothesis is corroborated by studies demonstrating a stronger impact of past positive outcome in

choices using either model-free or model-based measures [81,82,101].

Finally, it can be argued that if monkeys are a good model for human decision-making under uncertainty, they should display a description–experience gap. To our knowledge, so far only one study explicitly tackled this issue [102]. Monkeys were asked to make repeated choices between safe, and risky options, whose outcome probability was either learned by experience or described by the ratio between colours on a rectangle. Replicating previous findings in monkeys, and in discordance with the standard result in humans, Heilbronner and Hayden found that monkeys were risk-seekers in the description domain. However, consistent with the gap observed in humans, they also found that risk-seeking behaviour was higher for experience-based cues.

To summarize, the literature seems to suggest that monkeys’ decision-making for experience-based choice is quite consistent with what is observed in humans in terms of risk preference. This is consistent with a large body of literature showing that the neural substrates of reinforcement learning are largely preserved in the two species [103,104]. Risk seeking in this context may be driven by a higher learning rate from positive compared to negative prediction errors, which is essentially a computational reinforcement learning translation of the ‘hot hand’ fallacy [105,106]. The situation is much less reassuring concerning description-based decisions, as preferences compatible with PT are rarely observed. This can be due to the fact that pseudo description-based design in monkeys resembles the ‘description plus experience’ set-up in humans, where PT-like deformations are blunted or even disappear, as if description-based and experience-based biases reciprocally cancel themselves [78,79]. As a result, it remains unclear to what extent description-based processes can be elicited in the non-human primate animal model.

4. The impact of other experimental differences

Experimental results concerning decision-making under uncertainty in monkeys do not seem to straightforwardly comply with the predictions of PT. Overall it seems that monkeys’ behaviour is better accounted for as an experience-based decision process, which is consistent with the fact that pure description-based paradigms are not possible and monkey experiments always involve trial-by-trial feedback. The systematic presence of trial-by-trial feedback is not the only systematic methodological difference between the monkey and human studies (figures 2 and 4).

First, monkey studies essentially rely on primary rewards (mainly water or fruit juice), while human studies are realized

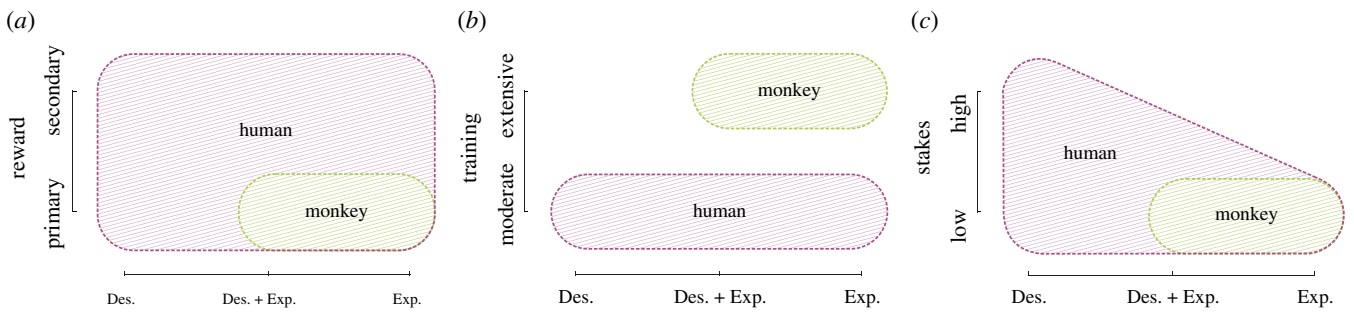


Figure 4. The figure illustrates how human (purple) and monkey (green) experimental settings map into a four-dimensional space, whose axes are: the way value information is provided (from description to experience); the nature of the reward (from primary to secondary; *a*), the amount of training (from moderate to extreme; *b*) and the level of the stakes (from low to high; *c*). (Online version in colour.)

Table 1. Studies investigating risk attitudes in rhesus monkeys. E, experience-based paradigms (i.e. without explicit representation of outcomes and probabilities); D, description-based paradigms (i.e. involving explicit representation of outcomes and probabilities; note that in monkeys this implies a 'description plus experience' set-up); liquid, the utilization of either water or fruit juice; tokens, the acquisition of a secondary reward, which is later exchanged for a primary reward; seek, an overall preference for the risky option; avoid, an overall preference for the safe option; inverse S-shape, the probability distortion postulated by prospect theory; S-shape, the probability distortion traditionally found in experience-based paradigms; N/A, the information is not available.

| study | sample size | modality | reward | risk attitude in gains | risk attitude in losses | probability distortion | loss aversion |
|---|-------------|----------|--------|------------------------|-------------------------|------------------------|---------------|
| McCoy & Platt [80] | 2 | E | liquid | seek | N/A | N/A | N/A |
| Hayden & Platt [81] | 2 | E | liquid | seek | N/A | N/A | N/A |
| Hayden <i>et al.</i> [82] | 5 | E | liquid | seek | N/A | N/A | N/A |
| Long [83] | 3 | E | liquid | seek | N/A | N/A | N/A |
| Watson [84] | 8 | E | liquid | seek | N/A | N/A | N/A |
| O'Neill & Schultz [85] | 2 | E | liquid | seek | N/A | N/A | N/A |
| Heilbronner <i>et al.</i> [86] | 3 | E | liquid | seek | N/A | N/A | N/A |
| Kim <i>et al.</i> [87] | 2 | E | liquid | seek | N/A | N/A | N/A |
| Heilbronner & Hayden [88] | 2 | E | liquid | seek | N/A | N/A | N/A |
| Xu & Kralik [89] | 2 | E | liquid | seek | N/A | N/A | N/A |
| Smith <i>et al.</i> [90] | 7 | E | liquid | seek | seek | N/A | N/A |
| Hayden <i>et al.</i> [91] | 4 | D | liquid | seek | N/A | N/A | N/A |
| So & Stuphorn [92] | 2 | D | liquid | seek | N/A | N/A | N/A |
| Yamada <i>et al.</i> [93] | ? | D | liquid | avoid | N/A | N/A | N/A |
| Raghuraman & Padoa-Schioppa [94] | 2 | D | liquid | seek | N/A | N/A | N/A |
| Staufer <i>et al.</i> [95] | 2 | D | liquid | seek | seek | inverse S-shape | N/A |
| Farashahi <i>et al.</i> [96], experiment 1 | 3 | D | liquid | seek | N/A | none | N/A |
| Farashahi <i>et al.</i> [96], experiment 2 | 3 | D | token | seek | seek | S-shape | N/A |
| Chen & Stuphorn [97] | 2 | D | liquid | seek | seek | inverse S-shape | N/A |
| Nioche <i>et al.</i> [98] | 2 | D | liquid | avoid | seek | inverse S-shape | yes |
| Ferrani-Toniolo <i>et al.</i> [99] experiment 1 | 2 | D | liquid | N/A | N/A | inverse S-shape | N/A |
| Ferrani-Toniolo <i>et al.</i> [99], experiment 2 | 2 | D | liquid | N/A | N/A | S-shape | N/A |
| Eisenreich <i>et al.</i> [100] | 3 | D | liquid | seek | seek | N/A | N/A |

mainly with secondary rewards (sometimes hypothetical ones) and primary reinforcers are only occasionally used [107,108]. Preliminary evidence from a study comparing risk propensity for different kinds of rewards in humans (money versus sport beverage) and monkeys showed similar patterns in the two species, thus suggesting that in more

comparable experimental condition risk preferences in both species could converge [109]. Furthermore, while the neural correlates of different kinds of rewards converge in the ventral prefrontal and striatal systems (principle of the common currency; [110]) they also have specific correlates, which may contribute to the different neural mechanisms

and result in distinct, reward-specific, risk preferences [107]. On the other side, a proxy for secondary reward can be found in monkey paradigms that involve collecting (virtual) tokens to be later exchanged for a primary reward. Unlike pure primary reward tasks, where losses cannot be implemented (it is impossible to take some fruit juice away from the stomach of a monkey), tokens have the advantage of making possible subtracting previously acquired rewards from the animal, thus inducing 'losses' in the same manner as in human. However, a recent study using tokens, showed risk-seeking attitudes comparable to that observed using primary reward [96]. Furthermore, when tokens are used, they are almost immediately changed against primary reward, making them not really comparable to money, whose value is much more permanent. Taken together, the available evidence suggests that the primary/secondary reward dichotomy does not explain the fact that human description-based biases are hardly observed in monkeys.

Second, in addition to the difference in the nature of the reward, description-based paradigms in humans and paradigms in monkeys often present a systematic difference in the amount of the reward (figure 4). Indeed, most of the original studies about PT used hypothetical gambles of hundreds of dollars and the same biases have been replicated using real stakes of about a month's salary [111]. On the other side, monkey studies use very small amounts of rewards (mere drops of liquids). It has been argued that part of the description–experience gap may simply derive from this difference in stake instead of being induced by fundamental differences in the decision-making process [88]. This would be consistent with Markowitz utility function which supposes risk seeking for small stakes (peanuts effect) before converting to risk aversion for higher stakes [112] and is supported by the finding that increasing the relative amount of reward (by reducing its frequency) decreases risk seeking down to risk neutrality in monkeys [88,112]. However, risk aversion in the gain domain (and a reverse pattern in the loss domain: the reflection effect) has also been observed with small stakes in description-based decisions in humans [67]. Thus, available evidence suggests that differences in the size of the stake cannot *fully* explain the fact that human description-based preferences are hardly observed in monkeys.

Finally, another notable difference between human and monkey experiments is represented by the amount and the type of training required to perform the task (see figures 2 and 4). In human experiments, task training rarely takes more than a few minutes (in some extreme cases of description-based paradigms, there is virtually no training: subjects are just *asked* to reveal their preferences). On the other side, monkey experiments require extensive training, in general spanning several months (usually training takes longer than the experiment itself). It can be, therefore, argued that their behaviour becomes to some extent habitual or automatized: a cognitive state that contrasts dramatically with the declarative and deliberative stance of description-based choices taken by humans [113]. In addition to that, training in monkeys (and other animals) often involves simplified versions of the task (often deterministic contingencies), which may reinforce specific risk preferences. Although the role of extended (several days, weeks) training and the resulting behavioural automation (or habituation) in risk preferences is unclear, it may contribute to the fact that human description-based biases are rarely observed in monkeys.

5. Conclusion and perspectives

Our review suggests that the rhesus monkey is a *partial* model of human decision-making under uncertainty. Risk preferences in monkeys are generally better explained as experience-based processes. Accordingly, monkeys proved to be a very good model of human reinforcement learning processes, providing crucial insights into its neural implementation (the dopamine prediction error hypothesis: [56,62,114]). The situation is less clear concerning description-based choices. In paradigms using explicit symbolic information about decision variables, monkeys only rarely displayed risk preferences compatible with human results. Deciding by description implies a symbolic system of communication. While in humans this system pre-exists (language), in monkeys it has to be learnt by trial-and-error, thus irremediably confounding description and experience. In addition to differences in the way value information is conveyed (experience- or description-based), other methodological factors (training, reward type and stakes) further drive apart the experimental set-ups of the two species. This situation is problematic as building a neural model of decision-making under uncertainty should integrate human (fMRI) and monkey (single unit) neurophysiological data, while explaining risk preferences in a wide range of situations that span from pure description-based choices to pure experience-based choices.

We propose further lines of research that could eventually help filling these gaps and ultimately fulfilling the strong neuroeconomic agenda. On the human side, the description–experience gap has been extensively studied at the behavioural level, but surprisingly neglected at the neural level. A notable exception [115], found different neural representations for description- and experience-oriented decisions. Furthering this line of enquiry would prove useful to redefine the target areas to look *specifically* for description-based processes in monkey electrophysiological studies.

With the development of online testing techniques, it is becoming easier to implement extended massive training in humans [116]. Translated in the field of decision-making under risk, these experiments would provide crucial insights into the impact of extensive training in risk preferences. While, description-based studies in monkeys require learning *ex novo* a symbolic system, in humans the meaning of pie-charts is provided by the language. It would be interesting to put humans in situations where they have to figure out by trial-and-error the code linking continuous visual features to decision variables.

In general, all the efforts aimed at increasing the methodological overlap between human and monkey studies will provide further insights into what are the behavioural processes shared across the two species. Popularizing fMRI experiments in monkeys would help confirm the neuro-anatomical targets and increase the focus on shared neural systems. The token paradigm (conceptually closer to the notion of the secondary reward) offers the possibility to implement losses in monkeys, hence facilitating the cross-species study of loss aversion.

Finally, on the monkey side, PT has been sporadically replicated. It will be important to clarify and formalize the experimental factors (in terms of stimuli, training and reward type; see table 1) that predict whether PT-like behaviour will be observed in a monkey experiment [88]. Determining under which experimental conditions PT is replicated in monkeys will imply a deeper understanding of the cognitive mechanisms underlying decision-making under uncertainty.

Data accessibility. This article has no additional data.

Authors' contributions. S.P. designed the review. B.G., F.C. and S.P. discussed the review. B.G., F.C. and S.P. wrote the review.

Competing interests. We declare we have no competing interests.

Funding. S.P. was supported by an ATIP-Avenir grant (R16069JS), the Programme Emergence(s) de la Ville de Paris, the Fyssen Foundation, the Fondation Schlumberger pour l'Éducation et la Recherche (FSER) and the CNRS projet 80 I Prime.

Endnotes

¹Subjective utility (or subjective value) representation seems to be distributed across a network of areas that include the ventral and the dorsal prefrontal cortices (both medial and lateral part), posterior cingulate cortex, the striatum, the insula, the amygdala and the hippocampus [26–28].

²It is indeed the case that brain systems encoding positive and negative values are, at least partially, dissociable. Losses are generally encoded by the insula, the amygdala and the dorsal prefrontal cortex, while gains are generally encoded in the ventral prefrontal and the striatum [31,32].

³Other non-invasive imaging techniques, such as magneto- and electro-encephalography present no advantage over fMRI when it comes to inferring single unit activity. They present better temporal resolution traded off against a worst spatial resolution.

⁴There are a few exceptions of single unit recordings in humans, obtained from neurologic patients undergoing brain surgery. While informative, these data are limited by the fact the neuro-anatomical targets cannot be chosen freely and that findings may not generalize to the general population [41].

⁵Of course, there is a lot of information to be gained also in the case where humans and monkeys do not display the same decisions and biases. Such differences currently represent a strong area of research in comparative psychology and ethology [44]. However, the (not so implicit) assumption of the vast majority of research in neuro-economics is that monkeys are valid experimental models for human cognition, and they are not investigated for comparative reasons.

⁶In the human reinforcement learning literature, the most frequently used paradigms involve options that possess, at a given trial, different expected values but overall similar risk level [56,62]. As a result the human reinforcement learning literature is more concerned about measures of objective performance rather than subjective preference.

References

- Samuelson PA. 1938 A note on the pure theory of consumer's behaviour. *Economica* **5**, 61–71. (doi:10.2307/2548836)
- Von Neumann J, Morgenstern O. 1944 *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
- Keynes JM. 1936 *The general theory of employment, interest, and money*. Berlin, Germany: Springer.
- Friedman M. 1953 *Essays in positive economics*. Chicago, IL: University of Chicago Press.
- Allais M. 1953 Le comportement de l'Homme rationnel devant le risque: critique des postulats et axiomes de l'Ecole Americaine. *Econometrica* **21**, 503–546. (doi:10.2307/1907921)
- Risk ED. 1961 Ambiguity, and the savage axioms. *Q. J. Econ.* **75**, 643–669. (doi:10.2307/1884324)
- Simon HA. 1955 A behavioral model of rational choice. *Q. J. Econ.* **69**, 99–118. (doi:10.2307/1884852)
- Tversky A, Kahneman D. 1979 Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–291. (doi:10.2307/1914185)
- Kahneman D, Slovic SP, Slovic P, Tversky A. 1982 *Judgment under uncertainty: heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Tversky A, Kahneman D. 1992 Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* **5**, 297–323. (doi:10.1007/BF00122574)
- Kahneman D, Knetsch JL, Thaler RH. 1991 Anomalies: the endowment effect, loss aversion, and status quo bias. *J. Econ. Perspect.* **5**, 193–206. (doi:10.1257/jep.5.1.193)
- Camerer CF. 1998 Prospect theory in the wild: evidence from the field. In *Advances in behavioral economics* (eds CF Camerer, G Loewenstein, M Rabin), pp. 148–161. Princeton, NJ: Princeton University Press.
- Vlaev I, Chater N, Stewart N, Brown GDA. 2011 Does the brain calculate value? *Trends Cogn. Sci.* **15**, 546–554. (doi:10.1016/j.tics.2011.09.008)
- Marr D, Poggio T. 1976 *From understanding computation to understanding neural circuitry*. Technical Report. Cambridge, MA: Massachusetts Institute of Technology.
- Friston KJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RS. 1995 Spatial registration and normalization of images. *Hum. Brain Mapp.* **3**, 165–189. (doi:10.1002/hbm.460030303)
- Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RS. 1994 Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* **2**, 189–210. (doi:10.1002/hbm.460020402)
- Worsley KJ, Friston KJ. 1995 Analysis of fMRI time-series revisited—again. *Neuroimage.* **23**, 173–181. (doi:10.1006/nimg.1995.1023)
- Nordhausen CT, Maynard EM, Normann RA. 1996 Single unit recording capabilities of a 100 microelectrode array. *Brain Res.* **726**, 129–140. (doi:10.1016/0006-8993(96)00321-6)
- Baker JT, Patel GH, Corbetta M, Snyder LH. 2006 Distribution of activity across the monkey cerebral cortical surface, thalamus and midbrain during rapid, visually guided saccades. *Cereb. Cortex.* **16**, 447–459. (doi:10.1093/cercor/bhi124)
- Camerer CF, Loewenstein G, Prelec D. 2004 Neuroeconomics: why economics needs brains. *Scand. J. Econ.* **106**, 555–579. (doi:10.1111/j.0347-0520.2004.00377.x)
- Camerer C, Loewenstein G, Prelec D. 2005 Neuroeconomics: how neuroscience can inform economics. *J. Econ. Lit.* **43**, 9–64. (doi:10.1257/0022051053737843)
- Rustichini A. 2009 Neuroeconomics: what have we found, and what should we search for. *Curr. Opin Neurobiol.* **19**, 672–677. (doi:10.1016/j.conb.2009.09.012)
- Chua HF, Gonzalez R, Taylor SF, Welsh RC, Liberzon I. 2009 Decision-related loss: regret and disappointment. *Neuroimage* **47**, 2031–2040. (doi:10.1016/j.neuroimage.2009.06.006)
- Coricelli G, Critchley HD, Joffily M, O'Doherty JP, Sirigu A, Dolan RJ. 2005 Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. Neurosci.* **8**, 1255–1262. (doi:10.1038/nn1514)
- De Martino B, Kumaran D, Seymour B, Dolan RJ. 2006 Frames, biases, and rational decision-making in the human brain. *Science* **313**, 684–687. (doi:10.1126/science.1128356)
- Lebreton M, Jorge S, Michel V, Thirion B, Pessiglione M. 2009 An automatic valuation system in the human brain: evidence from functional neuroimaging. *Neuron* **64**, 431–439. (doi:10.1016/j.neuron.2009.09.040)
- Bartra O, McGuire JT, Kable JW. 2013 The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **7**, 412–427. (doi:10.1016/j.neuroimage.2013.02.063)
- Clithero JA, Rangel A. 2014 Informatic parcellation of the network involved in the computation of subjective value. *Soc. Cogn. Affect. Neurosci.* **9**, 1289–1302. (doi:10.1093/scan/nst106)
- Palminteri S, Khamassi M, Joffily M, Coricelli G. 2015 Contextual modulation of value signals in reward and punishment learning. *Nat. Commun.* **6**, 1–14. (doi:10.1038/ncomms9096)
- Baillon A, Bleichrodt H, Spinu V. 2020 Searching for the reference point. *Manag. Sci.* **66**, 93–112. (doi:10.1287/mnsc.2018.3224)
- Palminteri S, Pessiglione M. 2017 Opponent brain systems for reward and punishment learning: causal evidence from drug and lesion studies in humans.

- In *Decision neuroscience* (eds J-C Dreher, L Tremblay), pp. 291–303. London, UK: Elsevier.
32. Pessiglione M, Delgado MR. 2015 The good, the bad and the brain: neural correlates of appetitive and aversive values underlying decision making. *Curr. Opin. Behav. Sci.* **5**, 78–84. (doi:10.1016/j.cobeha.2015.08.006)
 33. Tom SM, Fox CR, Trepel C, Poldrack RA. 2007 The neural basis of loss aversion in decision-making under risk. *Science* **315**, 515–518. (doi:10.1126/science.1134239)
 34. Bunge M. 2014 *The mind–body problem: a psychobiological approach*. Oxford, UK: Pergamon Press.
 35. Glimcher PW. 2011 *Foundations of neuroeconomic analysis*. Oxford, UK: Oxford University Press.
 36. Padoa-Schioppa C, Rustichini A. 2014 Rational attention and adaptive coding: a puzzle and a solution. *Am. Econ. Rev.* **104**, 507–513. (doi:10.1257/aer.104.5.507)
 37. Goense J, Bohraus Y, Logothetis NK. 2016 fMRI at high spatial resolution: implications for BOLD-models. *Front. Comput. Neurosci.* **10**, 66. (doi:10.3389/fncom.2016.00066/abstract)
 38. Iranpour J, Morrot G, Claise B, Jean B, Bonny J-M. 2015 Using high spatial resolution to improve BOLD fMRI detection at 3T. *PLoS ONE* **10**, e0141358. (doi:10.1371/journal.pone.0141358)
 39. Kayser C, Logothetis NK. 2010 The electrophysiological background of the fMRI signal. In *fMRI* (eds S Ulmer, O Jansen), pp. 23–33. Berlin, Germany: Springer.
 40. Logothetis NK. 2008 What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878. (doi:10.1038/nature06976)
 41. Zaghoul KA, Blanco JA, Weidemann CT, McGill K, Jaggi JL, Baltuch GH, Kahana MJ. 2009 Human substantia nigra neurons encode unexpected financial rewards. *Science* **323**, 1496–1499. (doi:10.1126/science.1167342)
 42. Jacobsen CF, Nissen HW. 1937 Studies of cerebral function in primates. IV. The effects of frontal lobe lesions on the delayed alternation habit in monkeys. *J. Comp. Psychol.* **23**, 101–112. (doi:10.1037/h0056632)
 43. Kennerley SW, Walton ME. 2011 Decision making and reward in frontal cortex: complementary evidence from neurophysiological and neuropsychological studies. *Behav. Neurosci.* **125**, 297. (doi:10.1037/a0023575)
 44. Addessi E, Beran MJ, Bourgeois-Gironde S, Brosnan SF, Leca J-B. 2020 Are the roots of human economic systems shared with non-human primates? *Neurosci. Biobehav. Rev.* **109**, 1–15. (doi:10.1016/j.neubiorev.2019.12.026)
 45. Fouragnan EF *et al.* 2019 The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. *Nat. Neurosci.* **22**, 797–808. (doi:10.1038/s41593-019-0375-6)
 46. Ruff CC, Driver J, Bestmann S. 2009 Combining TMS and fMRI: from ‘virtual lesions’ to functional-network accounts of cognition. *Cortex* **45**, 1043–1049. (doi:10.1016/j.cortex.2008.10.012)
 47. Gläscher J, Adolphs R, Damasio H, Bechara A, Rudrauf D, Calamia M, Paul LK, Tranel D. 2012 Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proc. Natl Acad. Sci. USA* **109**, 14 681–14 686. (doi:10.1073/pnas.1206608109)
 48. Si Y *et al.* 2018 Different decision-making responses occupy different brain networks for information processing: a study based on EEG and TMS. *Cerebral Cortex* **29**, 4119–4129. (doi:10.1093/cercor/bhy294)
 49. Goldstein WM, Hogarth RM. 1997 Judgment and decision research: some historical context. In *Research on judgment and decision making: currents, connections, and controversies* (eds WM Goldstein, RM Hogarth), pp. 3–65. Cambridge, UK: Cambridge University Press.
 50. Savage LJ. 1972 *The foundations of statistics*. New York, NY: Dover Publications Inc.
 51. Lichtenstein S, Slovic P. 2006 *The construction of preference*. Cambridge, UK: Cambridge University Press.
 52. Platt ML, Glimcher PW. 1999 Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238. (doi:10.1038/22268)
 53. Fiorillo CD, Tobler PN, Schultz W. 2003 Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299**, 1898–1902. (doi:10.1126/science.1077349)
 54. Wu G, Gonzalez R. 1996 Curvature of the probability weighting function. *Manag. Sci.* **42**, 1676–1690. (doi:10.1287/mnsc.42.12.1676)
 55. Hertwig R, Barron G, Weber EU, Erev I. 2004 Decisions from experience and the effect of rare events in risky choice. *Psychol. Sci.* **15**, 534–539. (doi:10.1111/j.0956-7976.2004.00715.x)
 56. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. 2006 Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**, 1042–1045. (doi:10.1038/nature05051)
 57. Fisher AJ, Medaglia JD, Jeronimus BF. 2018 Lack of group-to-individual generalizability is a threat to human subjects research. *Proc. Natl Acad. Sci. USA* **115**, E6106–E6115. (doi:10.1073/pnas.1711978115)
 58. Ludvig EA, Spetch ML. 2011 Of black swans and tossed coins: is the description–experience gap in risky choice limited to rare events? *PLoS ONE* **6**, e20262. (doi:10.1371/journal.pone.0020262)
 59. Thompson WR. 1933 On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* **25**, 285–294. (doi:10.1093/biomet/25.3-4.285)
 60. Slivkins A. 2019 Introduction to multi-armed bandits. *arXiv* 190407272.
 61. Sutton RS, Barto AG. 2018 *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press.
 62. Frank MJ, Seeberger LC, O’Reilly RC. 2004 By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science* **306**, 1940–1943. (doi:10.1126/science.1102941)
 63. Barron G, Erev I. 2003 Small feedback-based decisions and their limited correspondence to description-based decisions. *J. Behav. Decis. Mak.* **16**, 215–233. (doi:10.1002/bdm.443)
 64. Weber EU, Shafir S, Blais A-R. 2004 Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. *Psychol. Rev.* **111**, 430. (doi:10.1037/0033-295X.111.2.430)
 65. Hertwig R, Erev I. 2009 The description–experience gap in risky choice. *Trends Cogn. Sci.* **13**, 517–523. (doi:10.1016/j.tics.2009.09.004)
 66. Wulff DU, Mergenthaler-Canseco M, Hertwig R. 2018 A meta-analytic review of two modes of learning and the description–experience gap. *Psychol. Bull.* **144**, 140. (doi:10.1037/bul0000115)
 67. Madan CR, Ludvig EA, Spetch ML. 2019 Comparative inspiration: from puzzles with pigeons to novel discoveries with humans in risky choice. *Behav. Processes* **160**, 10–19. (doi:10.1016/j.beproc.2018.12.009)
 68. Abdellaoui M, Bleichrodt H, Kammoun H. 2013 Do financial professionals behave according to prospect theory? An experimental study. *Theory Decis.* **74**, 411–429. (doi:10.1007/s11238-011-9282-3)
 69. Prelec D. 1998 The probability weighting function. *Econometrica* **66**, 497. (doi:10.2307/2998573)
 70. Loomes G, Sugden R. 1982 Regret theory: an alternative theory of rational choice under uncertainty. *Econ. J.* **92**, 805–824. (doi:10.2307/2232669)
 71. Quiggin J. 2012 *Generalized expected utility theory: the rank-dependent model*. Dordrecht, The Netherlands: Springer.
 72. Lefebvre G, Lebreton M, Meyniel F, Bourgeois-Gironde S, Palminteri S. 2017 Behavioural and neural characterization of optimistic reinforcement learning. *Nat. Hum. Behav.* **1**, 1–9. (doi:10.1038/s41562-017-0067)
 73. Chambon V, Théro H, Vidal M, Vandendriessche H, Haggard P, Palminteri S. 2020 Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nat. Hum. Behav.* **4**, 1067–1079. (doi:10.1038/s41562-020-0919-5)
 74. Palminteri S, Lefebvre G, Kilford EJ, Blakemore S-J. 2017 Confirmation bias in human reinforcement learning: evidence from counterfactual feedback processing. *PLoS Comput. Biol.* **13**, e1005684. (doi:10.1371/journal.pcbi.1005684)
 75. Niv Y, Edlund JA, Dayan P, O’Doherty JP. 2012 Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *J. Neurosci.* **32**, 551–562. (doi:10.1523/JNEUROSCI.5498-10.2012)
 76. Tobler PN, Christopoulos GI, O’Doherty JP, Dolan RJ, Schultz W. 2008 Neuronal distortions of reward probability without choice. *J. Neurosci.* **28**, 11 703–11 711. (doi:10.1523/JNEUROSCI.2870-08.2008)
 77. Ludvig EA, Madan CR, McMillan N, Xu Y, Spetch ML. 2018 Living near the edge: how extreme outcomes and their neighbors drive risky choice. *J. Exp. Psychol. Gen.* **147**, 1905–1918. (doi:10.1037/xge0000414)
 78. Jessup RK, Bishara AJ, Bussemeyer JR. 2008 Feedback produces divergence from prospect theory in descriptive choice. *Psychol. Sci.* **19**, 1015–1022. (doi:10.1111/j.1467-9280.2008.02193.x)

79. Erev I, Ert E, Plonsky O, Cohen D, Cohen O. 2017 From anomalies to forecasts: toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychol. Rev.* **124**, 369. (doi:10.1037/rev0000062)
80. McCoy AN, Platt ML. 2005 Risk-sensitive neurons in macaque posterior cingulate cortex. *Nat. Neurosci.* **8**, 1220–1227. (doi:10.1038/nn1523)
81. Hayden BY, Platt ML. 2007 Temporal discounting predicts risk sensitivity in rhesus macaques. *Curr. Biol.* **17**, 49–53. (doi:10.1016/j.cub.2006.10.055)
82. Hayden BY, Heilbronner SR, Nair AC, Platt ML. 2008 Cognitive influences on risk-seeking by rhesus macaques. *Judgm. Decis. Mak.* **3**, 389.
83. Long AB, Kuhn CM, Platt ML. 2009 Serotonin shapes risky decision making in monkeys. *Soc. Cogn. Affect. Neurosci.* **4**, 346–356. (doi:10.1093/scan/nsp020)
84. Watson KK, Ghodasra JH, Platt ML. 2009 Serotonin transporter genotype modulates social reward and punishment in rhesus macaques. *PLoS ONE* **4**, e4156. (doi:10.1371/journal.pone.0004156)
85. O'Neill M, Schultz W. 2010 Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron* **68**, 789–800. (doi:10.1016/j.neuron.2010.09.031)
86. Heilbronner S, Hayden BY, Platt M. 2011 Decision salience signals in posterior cingulate cortex. *Front. Neurosci.* **5**, 55. (doi:10.3389/fnins.2011.00055)
87. Kim S, Bobeica I, Gamo NJ, Arnsten AF, Lee D. 2012 Effects of α -2A adrenergic receptor agonist on time and risk preference in primates. *Psychopharmacology* **219**, 363–375. (doi:10.1007/s00213-011-2520-0)
88. Heilbronner SR, Hayden BY. 2013 Contextual factors explain risk-seeking preferences in rhesus monkeys. *Front. Neurosci.* **7**, 7. (doi:10.3389/fnins.2013.00007/abstract)
89. Xu ER, Kralik JD. 2014 Risky business: rhesus monkeys exhibit persistent preferences for risky options. *Front. Psychol.* **5**, 258. (doi:10.3389/fpsyg.2014.00258/abstract)
90. Smith TR, Beran MJ, Young ME. 2017 Gambling in rhesus macaques (*Macaca mulatta*): the effect of cues signaling risky choice outcomes. *Learn. Behav.* **45**, 288–299. (doi:10.3758/s13420-017-0270-5)
91. Hayden B, Heilbronner S, Platt M. 2010 Ambiguity aversion in rhesus macaques. *Front. Neurosci.* **4**, 166. (doi:10.3389/fnins.2010.00166)
92. So N-Y, Stuphorn V. 2010 Supplementary eye field encodes option and action value for saccades with variable reward. *J. Neurophysiol.* **104**, 2634–2653. (doi:10.1152/jn.00430.2010)
93. Yamada H, Tymula A, Louie K, Glimcher PW. 2013 Thirst-dependent risk preferences in monkeys identify a primitive form of wealth. *Proc. Natl Acad. Sci. USA* **110**, 15 788–15 793. (doi:10.1073/pnas.1308718110)
94. Raghuraman AP, Padoa-Schioppa C. 2014 Integration of multiple determinants in the neuronal computation of economic values. *J. Neurosci.* **34**, 11 583–11 603. (doi:10.1523/JNEUROSCI.1235-14.2014)
95. Stauffer WR, Lak A, Bossaerts P, Schultz W. 2015 Economic choices reveal probability distortion in macaque monkeys. *J. Neurosci.* **35**, 3146–3154. (doi:10.1523/JNEUROSCI.3653-14.2015)
96. Farashahi S, Azab H, Hayden B, Soltani A. 2018 On the flexibility of basic risk attitudes in monkeys. *J. Neurosci.* **38**, 4383–4398. (doi:10.1523/JNEUROSCI.12260-17.2018)
97. Chen X, Stuphorn V. 2018 Inactivation of medial frontal cortex changes risk preference. *Curr. Biol.* **28**, 3114–3122.e4. (doi:10.1016/j.cub.2018.07.043)
98. Nioche A, Bourgeois-Gironde S, Boraud T. 2019 An asymmetry of treatment between lotteries involving gains and losses in rhesus monkeys. *Sci. Rep.* **9**, 1–13. (doi:10.1038/s41598-019-46975-2)
99. Ferrari-Toniolo S, Bujold PM, Schultz W. 2019 Probability distortion depends on choice sequence in rhesus monkeys. *J. Neurosci.* **39**, 2915–2929. (doi:10.1523/JNEUROSCI.1454-18.2018)
100. Eisenreich BR, Hayden BY, Zimmermann J 2019 Macaques are risk-averse in a freely moving foraging task. *Sci. Rep.* **9**, 15091. (doi:10.1038/s41598-019-51442-z)
101. Farashahi S, Donahue CH, Hayden BY, Lee D, Soltani A. 2019 Flexible combination of reward information across primates. *Nat. Hum. Behav.* **3**, 1215–1224. (doi:10.1038/s41562-019-0714-3)
102. Heilbronner SR, Hayden BY. 2016 The description-experience gap in risky choice in nonhuman primates. *Psychon. Bull. Rev.* **23**, 593–600. (doi:10.3758/s13423-015-0924-2)
103. Niv Y. 2009 Reinforcement learning in the brain. *J. Math. Psychol.* **53**, 139–154. (doi:10.1016/j.jmp.2008.12.005)
104. Daw ND, Doya K. 2006 The computational neurobiology of learning and reward. *Curr. Opin. Neurobiol.* **16**, 199–204. (doi:10.1016/j.conb.2006.03.006)
105. Croson R, Sundali J. 2005 The gambler's fallacy and the hot hand: empirical data from casinos. *J. Risk Uncertain.* **30**, 195–209. (doi:10.1007/s11166-005-1153-2)
106. Blanchard TC, Wilke A, Hayden BY. 2014 Hot-hand bias in rhesus monkeys. *J. Exp. Psychol. Anim. Learn. Cogn.* **40**, 280–286. (doi:10.1037/xan0000033)
107. Sescousse G, Caldú X, Segura B, Dreher J-C. 2013 Processing of primary and secondary rewards: a quantitative meta-analysis and review of human functional neuroimaging studies. *Neurosci. Biobehav. Rev.* **37**, 681–696. (doi:10.1016/j.neubiorev.2013.02.002)
108. Amiez C, Neveu R, Warrot D, Petrides M, Knoblach K, Procyk E. 2013 The location of feedback-related activity in the midcingulate cortex is predicted by local morphology. *J. Neurosci.* **33**, 2217–2228. (doi:10.1523/JNEUROSCI.2779-12.2013)
109. Hayden BY, Platt ML. 2009 Gambling for Gatorade: risk-sensitive decision making for fluid rewards in humans. *Anim. Cogn.* **12**, 201–207. (doi:10.1007/s10071-008-0186-8)
110. Levy DJ, Glimcher PW. 2012 The root of all value: a neural common currency for choice. *Curr. Opin. Neurobiol.* **22**, 1027–1038. (doi:10.1016/j.conb.2012.06.001)
111. Barberis NC. 2013 Thirty years of prospect theory in economics: a review and assessment. *J. Econ. Perspect.* **27**, 173–196. (doi:10.1257/jep.27.1.173)
112. Markowitz H. 1952 The utility of wealth. *J. Polit. Econ.* **60**, 151–158. (doi:10.1086/257177)
113. Wood W, Neal DT. 2007 A new look at habits and the habit-goal interface. *Psychol. Rev.* **114**, 843–863. (doi:10.1037/0033-295X.114.4.843)
114. Schultz W, Dayan P, Montague PR. 1997 A neural substrate of prediction and reward. *Science* **275**, 1593–1599. (doi:10.1126/science.275.5306.1593)
115. FitzGerald TH, Seymour B, Bach DR, Dolan RJ. 2010 Differentiable neural substrates for learned and described value and risk. *Curr. Biol.* **20**, 1823–1829. (doi:10.1016/j.cub.2010.08.048)
116. Wimmer GE, Poldrack RA. 2020 Reward learning and working memory: effects of massed versus spaced training and post-learning delay period. *bioRxiv*. 997098. (doi:10.1101/2020.03.19.997098)

5

The impassable gap between experiential and
symbolic values

1 **The impassable gap between experiential and symbolic values**

2 Basile Garcia (1), Maël Lebreton (2,3), Sacha Bourgeois-Gironde (4,5) &
3 Stefano Palminteri (1,6,§)

4 (1) Laboratoire de Neurosciences Cognitives Computationnelles, Département d'Etudes Cognitives, ENS,
5 PSL, INSERM, Paris, France;

6 (2) Paris School of Economics, Paris, France

7 (3) Swiss Center for Affective Science, Faculty of Psychology and Educational Sciences, University of
8 Geneva

9 (4) Institut Jean Nicod, Département d'Etudes Cognitives, ENS, EHESS, PSL, CNRS, Paris, France

10 (5) Assas CRED – Université Panthéon-Assas - Paris 2, Paris, France

11 (6) Institute of Cognitive Neuroscience, Higher School of Economics, Moscow, Federation of Russia

12 § corresponding authors (basile.garcia@ens.fr, stefano.palminteri@ens.fr)

13 **Abstract**

14 To choose between options of different natures, standard decision models presume that
15 a single representational system ultimately indexes their subjective values on a common
16 scale, regardless of how they are constructed. To challenge this assumption, we
17 systematically investigated hybrid decisions between experiential options, whose value is
18 built from past outcomes experience, and symbolic options which describe probabilistic
19 outcomes. We show that participants' choices exhibited a pattern consistent with a
20 systematic neglect of the experiential values. This normatively irrational decision strategy
21 held after accounting for alternative explanations, and persisted when it bore an economic
22 cost. Overall, our results demonstrate that experiential and symbolic values are not
23 symmetrically considered in hybrid decisions, suggesting that they are not
24 commensurable and recruit different representational systems which may be assigned
25 different priority levels in the decision process. These findings challenge the dominant
26 models commonly used in value-based decision-making research.

27 Introduction

28 Standard models of economic decision-making generally assume a two-step decision
29 process, where individuals identify and assign values to available options, and ultimately
30 pick the option with the highest subjective value (1–3). The values attributed to individual
31 options can derive from different sources. On the one hand, a priori neutral stimuli acquire
32 positive or negative experiential values after association with past outcomes (rewards and
33 punishments) (4–6). On the other hand, the explicit description of an option’s possible
34 outcomes and their probabilities are combined to form a subjective expected value (7–
35 10). Such explicit descriptions may take many different forms, including written language
36 (from simple vignettes to fully specified numerical variables), a symbolic code
37 communicating the decision variables (payoffs and probability) in an unambiguous
38 manner, or a combination of the two (11).

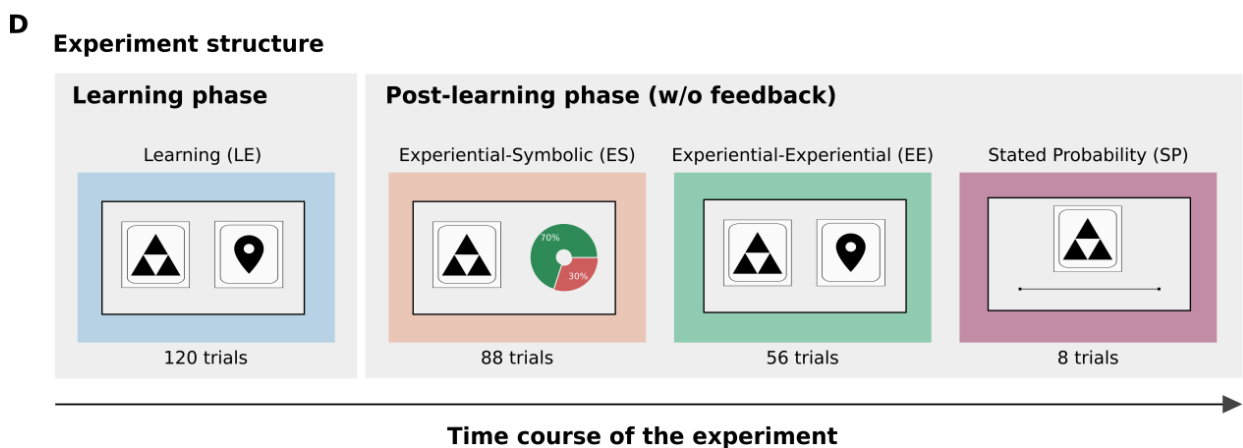
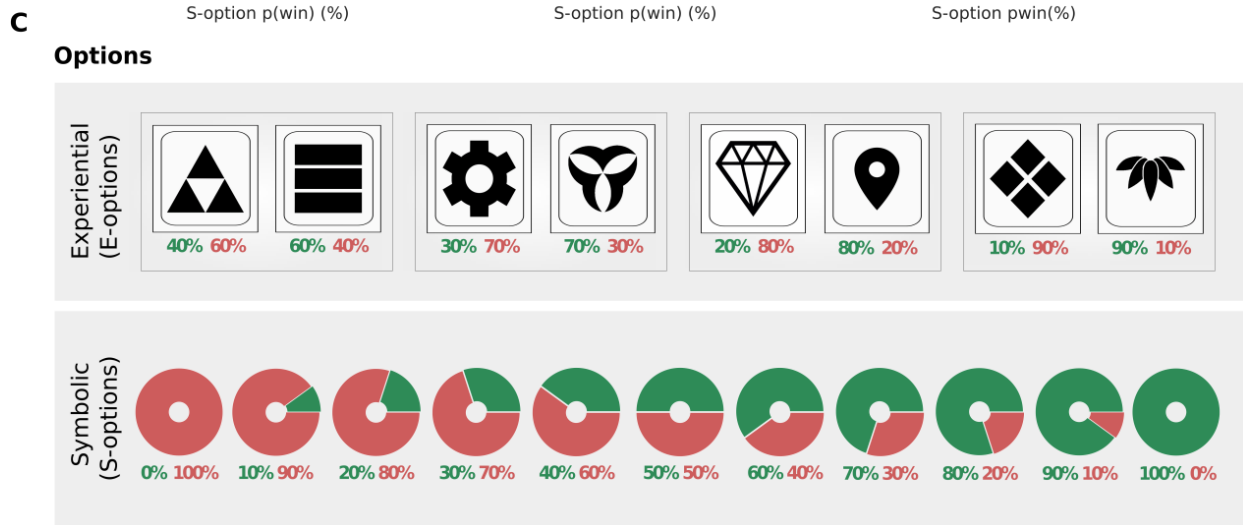
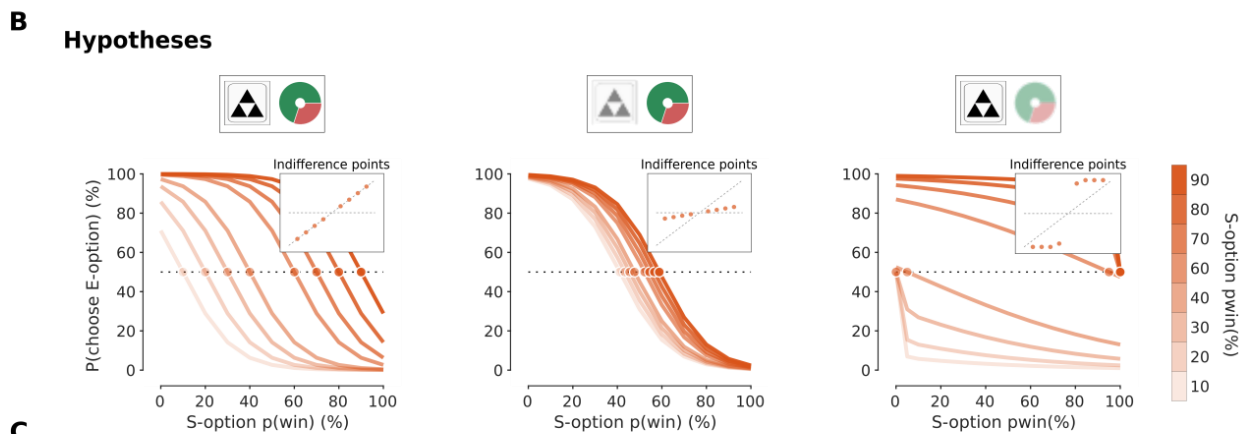
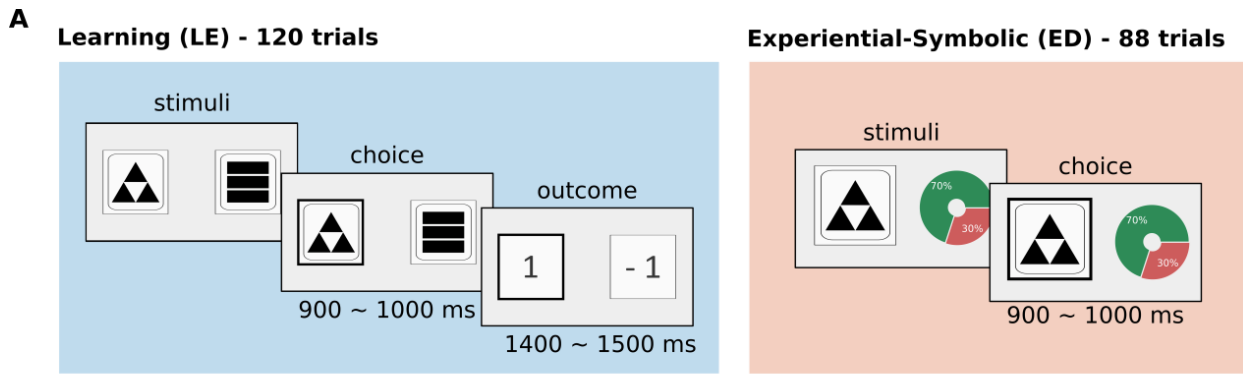
39
40 In the standard two-step model, the way option values are built (via experience or
41 description) is only peripheral to the decision process itself, meaning that experiential and
42 symbolic values converge to a central valuation and decision-making system (3, 12–16).
43 Thereby, choices *between* experiential and symbolic options should present no particular
44 challenge, because their values are translated into an internal common currency, allowing
45 an unbiased comparison between these differently generated option values. This
46 normative point of view is indirectly supported by the fact that the neural correlates of
47 experiential and symbolic values largely overlap in the so-called brain valuation system
48 (17–20).

49
50 However, several lines of evidence in behavioral decision-making research question the
51 idea of a central valuation system. In fact, it is now a very well established that, when
52 studied separately, experience-based and description-based choices display different
53 properties: a phenomenon referred to as the *description-experience gap* (21–24). This
54 difference in the subjective valuation of experiential and symbolic options poses a direct,
55 theoretical challenge to the idea of a central valuation system (25). This rather suggests
56 the existence of modality-specific valuation systems, relying on distinct cognitive

57 representations, which would hinder, if not impede, the comparison between experiential
58 and symbolic options.

59 Strikingly, this key prediction has not been directly assessed, because studies usually
60 consider separate sets of decision problems for experiential and symbolic options (26,
61 23). Thereby, to date, very little experimental evidence has formally assessed the
62 commensurability of experiential and symbolic option values, nor their mapping into a
63 central or different valuation systems (27, 28). This is particularly problematic considering
64 that hybrid choices seem to be the norm rather than the exception in our modern societies
65 where descriptive information is omnipresent. For example, everyday situations like
66 choosing between our favorite restaurant (experience) and a new one with good review
67 (description) is a prototypical example of such a hybrid decision.

68 To fill this gap and challenge the commensurability of experiential and symbolic values,
69 we designed a new behavioral protocol. The experiment started with a learning phase
70 during which human participants repeatedly faced abstract cues paired with probabilistic
71 outcomes, thereby learned to associate experiential expected-values to the originally
72 neutral symbols. After this phase, participants were asked to make hybrid choices
73 between the experienced symbols and described lotteries visualized as colored pie-charts
74 (a standard way to represent value symbolically) (11). When making hybrid choices,
75 participants treated the two kinds of options asymmetrically and, specifically, were
76 neglecting experiential values. This asymmetry was robust across seven experiments,
77 where we controlled for many possible alternative explanations, such as, insufficient
78 learning, generalization issues or lack of incentives. Overall, the relative neglect of an
79 option's value conditional on its source is consistent with the idea that different types of
80 values – such as experiential and symbolic – may involve different representational
81 systems, resulting in their incommensurability.



Time course of the experiment

83 **Fig 1. Behavioral tasks, hypotheses, option values and experimental protocol. (A)** The leftmost panel
 84 displays successive screens of a typical trials in the learning phase (LE). The LE-phase consists in a two-
 85 armed bandit task with fixed (4 or 2 – in Exp. 4) pairs of abstract cues (E-options) and contained 120 trials.
 86 The rightmost panel displays successive screens of a typical trials in the Experiential-Symbolic choice
 87 phase (ES). The ES-phase consists in binary choices between a lottery (standardly materialized as a pie-
 88 chart) and a symbol previously presented in LE-phase. In most experiments, the EE phase lasted 88 trials
 89 (8 E-options x 11 S-options). Durations are given in milliseconds. **(B)** The panels illustrate three possible
 90 hypotheses on how participants could make choices in the ES-phase. In each panel the probability of
 91 chosen the E-option is plotted against the value of the S-option (expressed as probability of winning a
 92 point). The insets represent the indifference points (where the curves cross 50%; of not unbiased
 93 indifference points should lay on the diagonal). The color of the curves indicates the value of the E-option
 94 (lowest: light orange; highest: dark orange). The leftmost panel illustrate the default hypotheses according
 95 to which E-options and S-options are fully commensurable and therefore the curves cross 50% (indifference
 96 point) at exactly the value of the E-option. The central panel illustrates *experiential value neglect* scenario
 97 according to which ES-choices are determined (almost) uniquely by the value of the S-options. Finally, the
 98 rightmost panel illustrates the symbolic value neglect scenario, accordingly to which ES-choices are
 99 determined (almost) uniquely the value of the E-options. **(C)** The panel displays the options values. The
 100 topmost part shows how E-option were organized in learning contexts (in all experiment except Exp. 4 and
 101 7; of note, the attribution of the value to the symbols was randomized across participants). The bottommost
 102 part shows the lotteries used in the ES phase (in all experiment except Exp. 7). **(D)** The experiments were
 103 structured as follows: they all started with a learning phase (LE), where participants made choices between
 104 abstract symbols and received feedback information. After the LE phase, participants were asked to make
 105 repeated choices between each E-option and several lotteries (see Fig. 1A and Fig. 1C). From Experiment
 106 5 on, participants were also asked to make choice between E-options that were not necessarily presented
 107 together. Finally, we assessed the stated probability (SP) of winning for each symbol by asking participants
 108 to explicitly rate each E-option, following a probability matching procedure (29).

109

110 Results

111 We conducted a series of experiments structured in two main phases, one allowing the
 112 formation of subjective values from the experience of past outcomes, and a second where
 113 these experiential options (E-options) were presented against options whose subjective
 114 values were described by symbolic means (S-options) (**Fig. 1A**). During the first (or
 115 learning: LE) phase, E-options were materialized by abstract shapes that provided no
 116 explicit information concerning the expected value (EV) of the option. During the LE
 117 choices, E-option values could therefore only be inferred from the history of gains (+1
 118 point) and losses (-1 point) associated to a specific cue. E-options were presented in four
 119 fixed pairs, each featuring an EV-maximizing and an EV-minimizing option.
 120 Subsequently, in the Experiential-Symbolic (ES) phase, participants were asked to make
 121 choices between the very same E-options of the previous phase and pie-charts explicitly
 122 describing the associated probabilities of gain and loss. As these ES, “hybrid” choices are

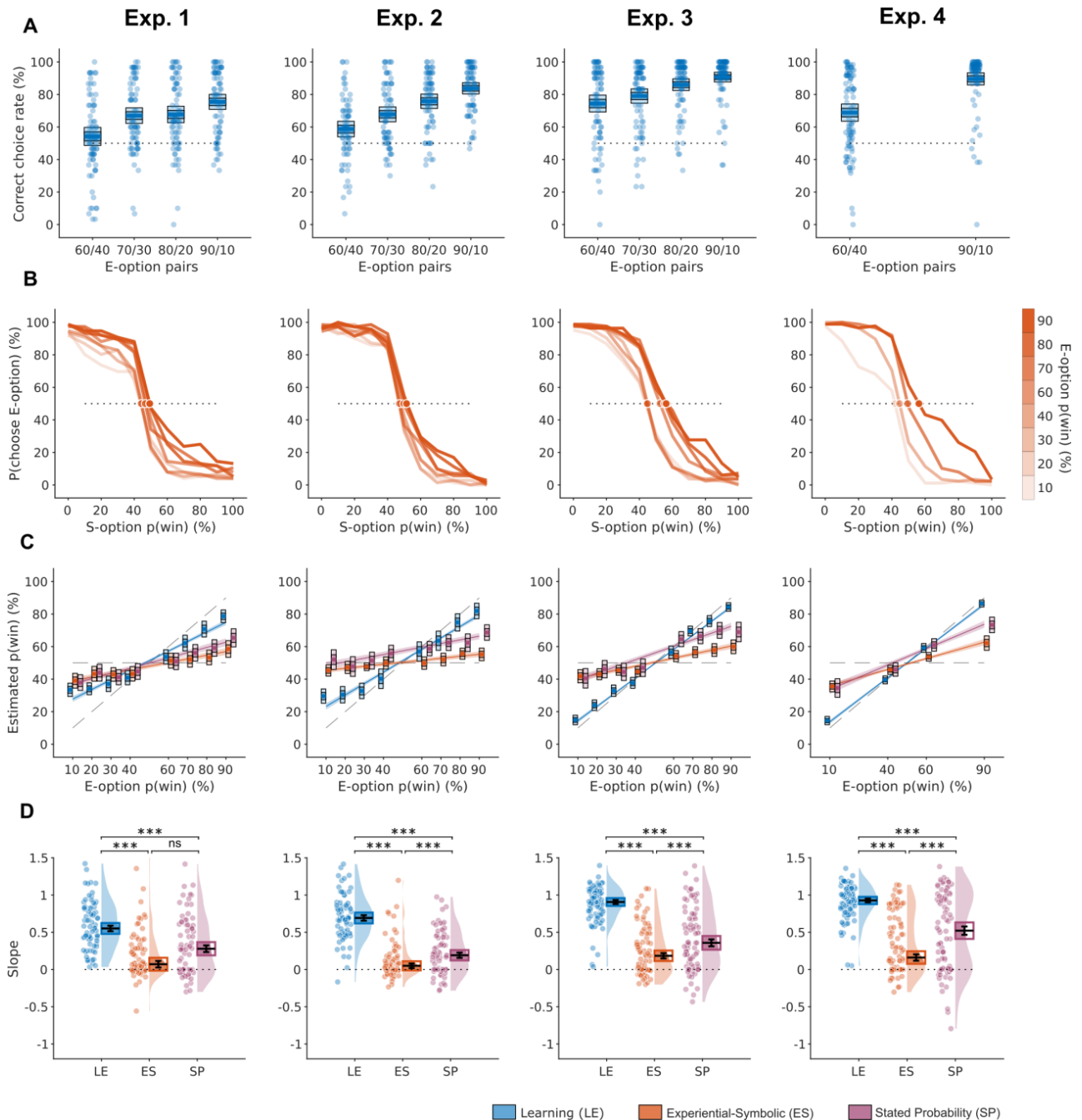
123 the main focus of this paper, we thereafter delineate three plausible hypotheses
124 concerning the behavioral output of this phase.

125 First, assuming that the subjective values of the E- and S-options are mapped into a
126 common scale (*common currency* hypothesis), participants should make *unbiased*
127 decisions in the ES phase. Accordingly, the probability of choosing, say, the E-option, will
128 be jointly determined by the EV of the E- and the S-option (**Fig. 1B**: left). In other terms,
129 for a given E-option the inferred *indifference point* will precisely correspond to S-options
130 with equal EV.

131 Alternatively, the possibility that subjective values are constructed and represented in a
132 modality-specific way (*representational gap* hypothesis) entails that E- and S-options are
133 not readily commensurable. This situation could lead to two possible scenarios. In one
134 of them, participants make random choices in the ES-phase. In the other scenario
135 participants could prioritize one of the two sources of information. Within this scenario,
136 participants could resolve the tension between E- and S-options basing their choices
137 primarily on the explicit symbolic values provided by the lotteries. In other terms,
138 participants would pick the lottery, when positive, and reject it when negative, as if the E-
139 option values were neglected and regressed to zero (*experiential value neglect*, **Fig. 1B**:
140 mid). In the other case, participants would present an over-reliance on experiential values
141 and would display the opposite pattern: accept or reject an E-option without considering
142 the S-option value (*symbolic value neglect*, **Fig. 1B**: right). Crucially, the ES phase of our
143 experiments allows to tease apart these different scenarios by analyzing the probability
144 of choosing an E-option as a function of the S-option being presented. More precisely,
145 taking each E-option separately and uncovering the S-option (value) at which a
146 preference shifts from the former to the latter provides us with an estimate of how much
147 a participant values an E-option. Quantifying the relation between E-options and S-
148 options boils down to inferring indifference points (i.e., when the probability of choosing

149 one option over the other is 50%) which acts as proxies of participant E-option values
 150 (Fig. 1B: insets).

151



152

153 **Fig 2. Raw behavioral results and inferred option values in Experiments 1-to-4.** (A) Correct choice
 154 rate grouped per learning context in the LE phase, where '40/60' designated the hardest decision problem,
 155 '10/90' the easiest decision problem. The dark blue line indicates the mean, the mid-dark blue indicates the
 156 standard mean error, and the light blue indicates a 95% confidence interval. The dotted line indicates
 157 chance (or random) responding (50%). (B) Average probability of choosing an E-option over a S-option
 158 during ES phase. The color of the curves indicates the value of the E-option (lowest: light orange; highest:

159 dark orange). Dots represent the empirical indifference points, the value of a lottery that corresponds to a
 160 probability of choosing the symbol 50% of the times. **(C)** The panels represent for each symbol the inferred
 161 value (as expressed by the probability of winning; $p(\text{win})$) as a function of the actual value. ES estimates
 162 are represented in orange, LE estimates in blue and SP estimates in pink. In the data-boxes, the dark tone
 163 line represents the mean, mid-dark tone the standard mean error, light tone a 95% confidence interval. The
 164 lines represent linear regression (dark tone), and the average standard mean error (light tone). **(D)**
 165 Comparison of individual inferred slopes obtained from linear fit (see **Fig. 2C**) in the three modalities (LE,
 166 ES and SP in blue, orange and pink, respectively). The black lines represent mean and standard error of
 167 the mean. The colored boxes represent 95% confidence interval. The shaded area probability represents
 168 density functions. *** $p < 0.001$ paired sample t-tests.

169

170 **First evidence for the experiential value neglect scenario**

171 In the LE phase of the first experiment (N=76), we presented pairs of E-options in an
 172 *interleaved* manner (i.e., E-option pairs are distributed randomly in the sequence of trials)
 173 and we displayed only the outcome of the chosen option (*partial feedback*) (**Fig. 2A, Exp.**
 174 **1**). Apart from the most difficult learning context (60/40), choice accuracy was above
 175 chance level for all E-option pairs ($T(75)=1.5$, $P > .05$; $T(75)=10.98$, $P < 0.001$), thus
 176 indicating that participants aimed at (and managed to) maximize expected value.
 177 Furthermore, accuracy was modulated by the difference in expected value (i.e., the
 178 *decision value*) of the E-option pair. Choice accuracy increased as a function of the
 179 decision value ($\beta=0.077$, $T(300)=2.16$, $P < 0.05$; $\beta=0.08$, $T(300)=2.35$, $P < 0.05$; $\beta=0.21$,
 180 $T(300)=5.94$, $P < 0.001$), thus indicating that participants' behavior was sensitive to the
 181 specific EV of E-options involved in a given pair.

182 Regarding analysis of the ES phase, the probability of choosing an E-option in an ES
 183 decision was largely determined by the S-option EV-value and the preference shift
 184 abruptly occurred around S-option EV equal to zero (i.e., $P(+1) = P(-1) = 0.5$). Despite
 185 clear proofs of successful value learning and encoding during the LE phase, ES phase-
 186 choice pattern was clearly consistent with the *experiential value neglect* scenario. (**Fig.**
 187 **2B: left**).

188 To quantify and statistically compare the differences in preferences observed in the LE
 189 and the ES phase, we first estimated the theoretical subjective value of each E-option
 190 separately for the two choice types, proxied by its probability of winning a point: $p(\text{win})$
 191 (remind that the outcomes are fixed, so the expected value of different options only
 192 depend on their probabilities to win). Concerning the LE phase, we leveraged on a

193 classical associative learning approach, where we assumed $p(\text{win})$ to be iteratively
194 updated as a function of a prediction error-minimizing learning rule (30, 31, 6). We were
195 able to infer $p(\text{win})$ attributed to each E-option at the end of the learning process by fitting
196 this, rather parsimonious and standard, model.

197 Concerning the ES phase, subjective $p(\text{win})$ estimates were inferred using the following
198 method: the probability of choosing a specific E-option over a S-option of various
199 expected values was assumed to take the form of a logistic sigmoid function. We fitted
200 those logistic functions to each E-option and individual, and used them to extrapolate the
201 indifference points indexing E-options' subjective $p(\text{win})$.

202 Finally, to compare the overall valuation of the E-options in the LE and the ES phases,
203 we computed a measure of how well the subjective $p(\text{win})$ estimates from each phase
204 matched the objective underlying probabilities, using slopes estimates from linear
205 regressions.

206 At this aggregate level, a slope equal to 1 corresponds to an unbiased representation of
207 E-options' $p(\text{win})$, whereas a slope equal to 0 corresponds to random representations. In
208 our data, the slopes estimated from the LE phase were significantly higher and closer to
209 1 compared to those estimated from ES-choices ($T(75)=6.53$, $P < .001$) (**Fig. 2C**: left).
210 Thus, ES decision problems feature a specific neglect of E-option values, as if hybrid
211 choices prioritized the value of the symbolic options over an unbiased comparison of
212 experiential and symbolic values, thereby confirming the *experiential neglect* hypothesis.

213 We ruled out a first trivial interpretation for this result, by only including in the analyses
214 participants that performed at 100% of correct response in catch trials (i.e. trials involving
215 choices between two S-options; see **Supplementary Materials**), disseminated across
216 the ES phase to ensure the participants' capacity to understand the symbolic
217 representation of the probabilities.

218 In the following sections of the paper, we provide additional evidence in favor of the
219 experiential neglect hypothesis by progressively ruling out alternative interpretations via
220 additional measures and experiments.

221

222 Ruling out insufficient learning and forgetting

223 While the *experiential neglect* pattern observed in the ES phase is consistent with the
224 idea that E-options and S-options are not equally considered in the decision process, it is
225 also consistent with a much more trivial hypothesis: insufficient learning. Despite
226 reinforcement learning model fitting suggesting otherwise (see **Fig. 2C**: left), it is indeed
227 possible that the neglect of E-option in the decision is caused by an imperfect and noisy
228 E-option value representations at the end of the learning phase. To rule out this alternative
229 interpretation, we devised a series of experiments where we changed the LE phase in
230 order to improve learning, while keeping the (average) option values the same. In a
231 second experiment (Exp. 2; N=71), we therefore presented decision problems as blocks
232 (rather than interleaved as in Exp. 1), so as to improve performance and option
233 identification by preventing the saturation of working memory (32). In a third experiment
234 (Exp. 3; N=83), we additionally provided the outcome information concerning the
235 unchosen option – a manipulation known for increasing accuracy (33, 34). Finally, on top
236 of these variations, in a fourth experiment (Exp. 4; N=88) we also reduced the number of
237 decision problems of the LE phase to two, such that each decision problem was presented
238 for twice as many trials as in experiments 1-3, thereby reducing the uncertainty about the
239 options' outcomes. These manipulations were successful in significantly increasing
240 decision accuracy in the LE phase (**Exp. 1**: 0.66 ± 0.01 ; **Exp. 2**: 0.71 ± 0.01 , $\beta=0.05$,
241 $T(314)=2.28$, $P < 0.05$; **Exp. 3**: 0.82 ± 0.01 , $\beta=0.16$, $T(314)=7.17$, $P < 0.001$; **Exp. 4**:
242 0.79 ± 0.01 ; $\beta=0.13$, $T(314)=5.8$, $P < 0.001$), while avoiding ceiling performance issues.
243 Indeed, even in the easiest experiments, accuracy was still significantly modulated by the
244 decision values; for instance, the accuracy in the more difficult decision problem (60/40)
245 was always lower compared to the easiest one ('90/10') ($T=5.81$, $P < 0.001$; $T=8.81$,
246 $P < 0.001$).

247 Crucially, the remarkable increase in the LE phase accuracy of the new experiments
248 (107% - 124% of Exp. 1) was not paralleled by detectable qualitative differences in ES
249 phase choice patterns (**Fig 2B**). In other terms, the *experiential value neglect* persists
250 despite the uncertainty concerning the E-options' values being considerably reduced (via

251 blocked design, complete feedback and increasing the number of trials per decision
252 problem).

253 To quantitatively characterize this claim, we estimated the subjective $p(\text{win})$ for each E-
254 option separately for the LE and the ES phases and fitted a linear regression between the
255 estimated subjective $p(\text{win})$ and their true values (as described above). Confirming the
256 efficiency of our manipulations in increasing learning performance, the LE-inferred slopes
257 increased significantly across experiments (**Exp. 2:** $\beta=0.11$, $T(942)=5.98$, $P=0.055$; **Exp.**
258 **3:** $\beta=0.28$, $T(942)=6.5$, $P < 0.001$; ; **Exp. 4:** $\beta=0.31$, $T(942)=7.27$, $P < 0.001$). Critically,
259 the ES slopes were not modulated across experiments aside from Exp. 4 (**Exp. 2:** $\beta=-$
260 0.1 , $T(942)=-1.76$, $P=0.07$; **Exp. 3:** $\beta=0.02$, $T(942)=6.5$, $P=0.67$; ; **Exp. 4:** $\beta=0.11$,
261 $T(942)=2.06$, $P < 0.05$) (**Fig. 2D**). Overall, LE-inferred slopes were significantly higher
262 than the ES slopes in all experiments (**Exp. 2:** $T(70)=11.74$, $P < 0.001$; **Exp. 3:**
263 $T(82)=15.8$, $P < 0.001$; **Exp. 4:** $T(87)=11.64$, $P < 0.001$; **Fig. 2E**), and the asymmetric
264 effects of the manipulations on the LE versus ES phases translated into a significant
265 interaction between the choice modality (ES and LE) and the experiment number (**Exp.**
266 **2:** $\beta=-0.21$, $T(942)=-2.58$, $P<0.05$; **Exp. 3:** $\beta=-0.26$, $T(942)=-3.29$, $P<0.01$; ; **Exp. 4:**
267 $\beta=0.2$, $T(942)=2.57$, $P < 0.05$).

268 The comparison between the first four experiments suggests that *experiential value*
269 *neglect* is not a mere effect of insufficient learning. We indeed observe that an improved
270 performance in the learning phase does not translate into a similar decrease of the
271 *experiential value neglect* effect. However, independently of the quality of learning, it is
272 also theoretically possible that participants forgot the E-option values when entering the
273 ES hybrid choice phase, although the fact that the ES phase directly succeeded the LE
274 phases within a matter of seconds makes it improbable. To rule out this possibility, in
275 Exp. 1-4, we asked participants to evaluate the E-options' $p(\text{win})$ just after the ES phase,
276 by implementing a fully incentivized stated probability (SP) procedure (35). More
277 precisely, participants were explicitly asked to rate the probability of winning a point they
278 attribute to an E-option, by means of a numerical rating scale (**Fig. 1D**).

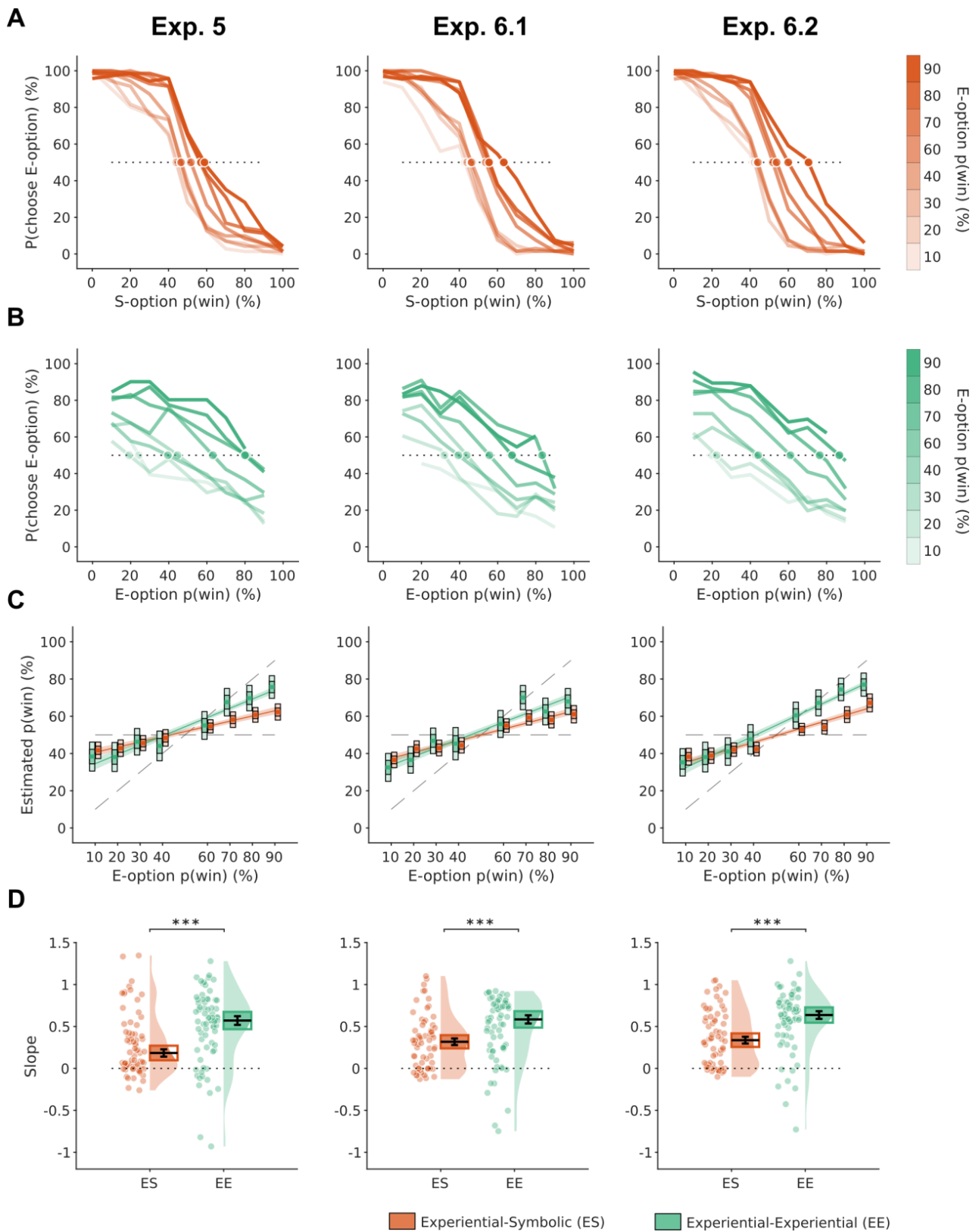
279 We then evaluated the quality of the E-option memory retention by regression these
280 stated probabilities against their true values. Note that because this elicitation happens

281 after the ES phase, this SP-inferred slopes constitutes a lower bound of how well E-option
282 values are learned and could be recovered during the ES phase. Yet, the SP-inferred
283 slopes were systematically higher than the ES-inferred slopes and significantly so in Exp.
284 2, 3, 4 (**Exp. 1:** $T(75)=2.62$, $P>0.05$; **Exp. 2:** $T(70)=3.42$, $P<0.05$; **Exp. 3:** $T(82)=4.38$,
285 $P<0.001$, **Exp. 4:** $T(87)=4.87$, $P<0.001$). Therefore, E-options' values elicited during the
286 SP phase were more accurate than those elicited in the preceding ES-phase. This
287 observation rules out forgetting as a plausible interpretation of the apparent *experiential*
288 *value neglect* pattern observed in the ES phase.

289 **Ruling out generalization issues and assessing the robustness to practice**

290
291 The above-reported results from 4 experiments and 3 preference elicitation methods
292 indicate that the *experiential value neglect* phenomenon cannot be accounted for by
293 insufficient learning nor by mere forgetting. In the present section we rule out two
294 additional alternative explanations. First, it should be noted that the ES phase involves a
295 generalization process, because the E-options are extrapolated from the decision context
296 where their subjective values are originally learned. It is therefore conceivable that the
297 apparent *experiential value neglect* is spuriously created by a generalization problem.
298 Second, in the previously reported experiments, participants went through the different
299 phases (LE, ES and SP) only once: perhaps participants were somehow taken by surprise
300 by the ES phase. In that case, presenting them different phases of the experiment twice
301 will possibly allow them to improve their decisions by anticipating the ES-phase (36).

302 To control for generalization and practice, we run two additional experiments. In Exp. 5
303 and Exp. 6 ($N=71$ and $N=66$), after the learning phase, we interleaved the ES-choices
304 with choices involving E-options presented in all possible combinations (referred to as
305 EE-choices). Thus, in all cases except one, EE-choices required being able to generalize
306 their value to new decision problems. As in ES-choices, we plotted the probability of
307 choosing a given E-option as a function of the alternative E-option (**Fig. 3B**). To check
308 whether *experiential value neglect* disappears if participants are given the opportunity to
309 learn how to make ES decisions, Exp. 6 included a second session where we repeated
310 all phases (LE, ES, ES and SP). Of note, E-options in the second sessions were
311 materialized by a new set of symbols.



312

313

314 **Fig 3. Raw behavioral results and inferred option values in Experiments 5-to-6. (A)** Average
 315 probability of choosing an E-option over a S-option during ES phase. The color of the curves indicates the
 316 value of the E-option (lowest: light orange; highest: dark orange). Dots represent the empirical indifference
 317 points, the value of a lottery that correspond to a probability of choosing the symbol 50% of the times. Exp.
 318 6.1 and Exp. 6.2 refers to the first and the second session, respectively. **(B)** Average probability of choosing
 319 an E-option over another E-option during EE phase. The color of the curves indicates the value of the E-
 320 option (lowest: light green; highest: dark green). Dots represent the empirical indifference points, the value
 321 of a lottery that corresponds to a probability of choosing the symbol 50% of the times. **(C)** The panels
 322 represent for each symbol the inferred value (as expressed by the probability of winning; $p(\text{win})$) as a
 323 function of the actual value. ES estimates are represented in orange and EE estimates in green. In the
 324 data-boxes, the dark tone line represents the mean, mid-dark tone the standard mean error, light tone a
 325 95% confidence interval. The lines represent linear regression (dark tone), and the average standard mean
 326 error (light tone). **(D)** Comparison of individual inferred slopes obtained from linear fit (see **Fig. 3C**) in two
 327 modalities (ES and EE in orange and green, respectively). The black lines represent mean and standard
 328 error of the mean. The colored boxes represent 95% confidence interval. The shaded area represents
 329 probability density functions. *** $p < 0.001$ two sample t-test.

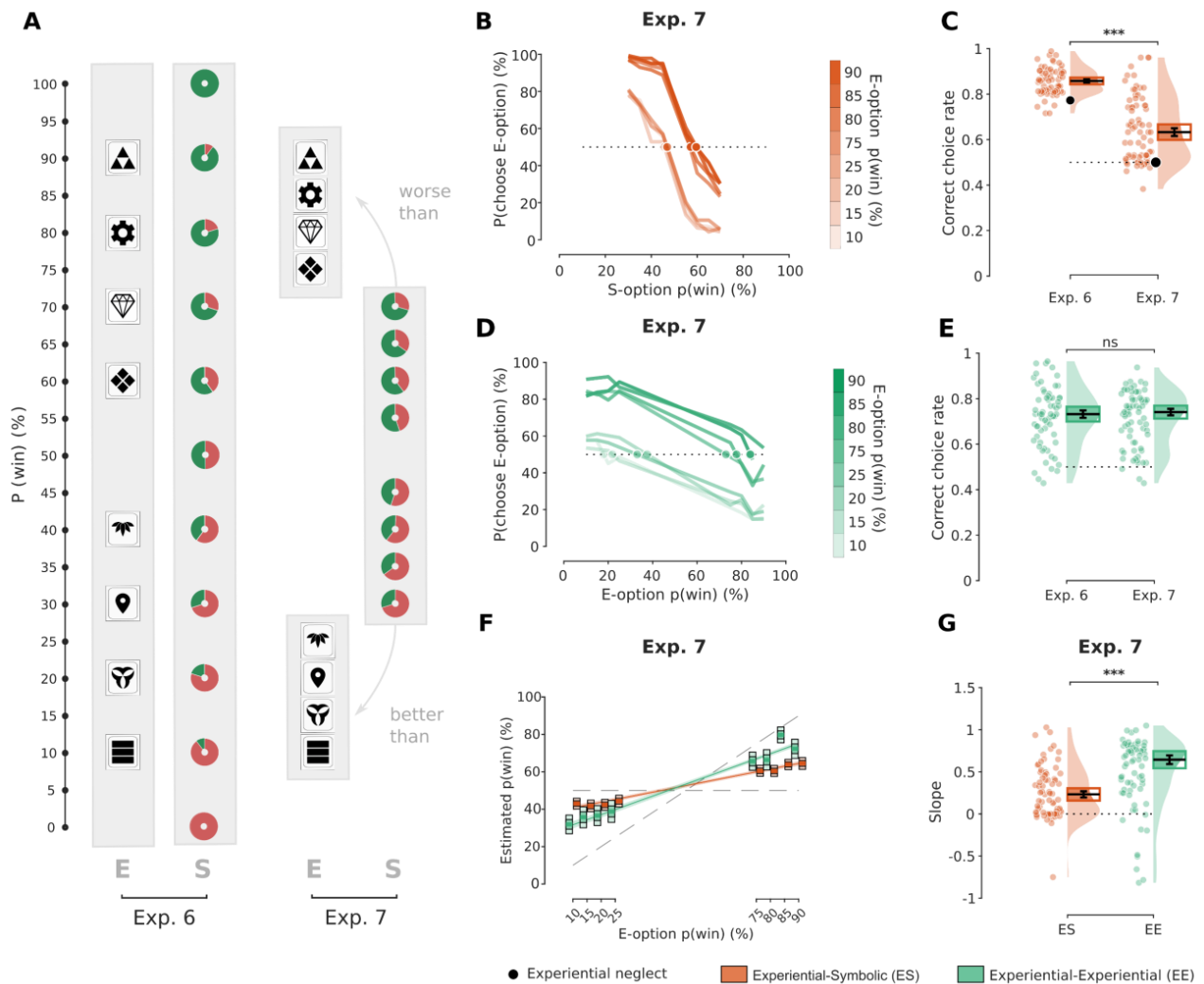
330
 331 EE-choices curves revealed that participants were capable of successfully extrapolating
 332 the value of the E-options to new decision problems involving other E-options. On the
 333 other side, the ES-choices were consistent with experiential values neglect, thus
 334 replicating the previous experiments (of note, the LE-phase of Exp. 5 and Exp. 6
 335 presented the same characteristics as that of Exp. 3: complete feedback and block
 336 design) (**Fig. 3A**).

337 To formally assess the difference between EE- and ES-choices, we calculated for each
 338 participant their option-specific indifference points, following the same procedure used for
 339 ES-choices and we compared the inferred slopes across decision modalities. EE-inferred
 340 slopes were consistently significantly higher than ES slopes in both Exp. 5 and Exp. 6
 341 (**Exp. 5:** $T(70)=4.5$, $P < 0.001$; **Exp. 6.1:** $T(65)=4.08$, $P < 0.001$).

342 Being presented with the whole experiment a second time had no detectable effect in
 343 choice behavior in neither the EE- or the ES-phase. Indeed, we observe no significant
 344 increase in the slopes in neither ES- ($\beta=0.04$, $T(260)=0.84$, $P=0.4$) nor EE- choices
 345 ($\beta=0.1$, $T(260)=1.59$, $P=0.11$) and the ES-inferred slopes were still significantly smaller
 346 compared to EE- ones (Exp. 6.2: $T(65)=5$, $P < 0.001$). This suggests that being exposed
 347 with the whole experiment one time and, by doing so giving participants the possibility to
 348 adjust the decision strategy does not affect the main results.

349

350



351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

Fig 4. Option values and behavioral results in Experiment 7. (A) The panel shows and compares the options value in Exp. 1-6 to that of Exp. 7. In Exp. 7, we reorganized E-options and S-options values such that half of the E-options have higher expected-values than all S-options and, conversely the other half have lower expected-values. In such an arrangement, a participant fully neglecting the E-options values in the ES phase will end up with random choices in respect to utility maximization **(B)** Average probability of choosing an E-option over a S-option during ES phase. The color of the curves indicates the value of the E-option (lowest: light orange; highest: dark orange). Dots represent the empirical indifference points, the value of a lottery that correspond to a probability of choosing the symbol 50% of the times. **(C)** Expected value maximizing (i.e., correct) choices in the ES phase of Exp. 6 compared to Exp. 7. The black lines represent mean and standard error of the mean. The colored boxes represent 95% confidence interval. The shaded area probability represents density functions. *** $p < 0.001$ two-sample t-test. **(D)** Average probability of choosing an E-option over another E-option during EE phase. The color of the curves indicates the value of the E-option (lowest: light green; highest: dark green). Dots represent the empirical indifference points, the value of a lottery that correspond to a probability of choosing the symbol 50% of the times. **(E)** Expected value maximizing (i.e., correct) choices in the EE phase of Exp. 6 compared to Exp. 7. The black lines represent mean and standard error of the mean. The colored boxes represent 95% confidence interval. The shaded area probability density functions. *** $p < 0.001$ two-sample t-test. **(F)** The panel

369 represents for each symbol the inferred value (as expressed by the probability of winning; $p(\text{win})$) as a
370 function of the actual value. ES estimates are represented in orange and EE estimates in green. In the
371 data-boxes, the dark tone line represents the mean, mid-dark tone the standard mean error, light tone a
372 95% confidence interval. The lines represent linear regression (dark tone), and the average standard mean
373 error (light tone). **(G)** Comparison of individual inferred slopes obtained from linear fit (see **Fig. 4F**) in two
374 modalities (ES and EE; in orange and green, respectively). The black lines represent mean and standard
375 error of the mean. The colored boxes represent 95% confidence interval. The shaded area probability
376 represents density functions. *** $p < 0.001$ paired two-sample t-test.

377

378 **Experiential value neglect persists even when it bears an economic cost**

379 Analysis of choice behavior in the ES show that learned values of the E-options are largely
380 neglected, as if participants were deciding on the basis of the value of the S-options only,
381 and this despite the fact performance in the LE, SP and EE-choices indicate that E-option
382 values are well learned and memorized. Neglecting experiential values seems, at least
383 *prima facie*, suboptimal for the decision process, as taking into account all relevant
384 information is considered a hallmark of normative behavior (37, 38). However, if E-option
385 information processing (e.g. memory access/retrieval) is costly or if neglecting E-options
386 does not hinders decision performance dramatically, it may become rational to do so (39–
387 41).

388 To evaluate this possibility, we simulated choices based on an extreme version of the
389 experiential neglect rule: if an S-option has positive expected value, choose it, otherwise
390 choose the E-option. These simulations show that, applied to the decision problems of
391 the ES phase from experiments 1-to-6, extreme experiential neglect still generates 77%
392 of expected-value maximizing choices. This result is actually not as counterintuitive as it
393 initially appears: by design, a positive lottery is the most advantageous option in $\geq 50\%$
394 of the decision problems in which it appears, and the converse is true for the negative
395 expected value lotteries. These considerations suggest that, instead of representing an
396 intrinsic cognitive limitation of value-based decision-making, the *experiential value*
397 *neglect* is a rational heuristic strategy deployed by efficient (or lazy) decision-makers
398 maximizing an accuracy-effort trade-off (42–45).

399 In order to test this new interpretation of the results, we designed a new experiment (Exp.
400 7) where we reorganized E- and S-options probabilities in a way that makes neglecting
401 experiential values economically disadvantageous (**Fig. 4A**). In this new configuration,

402 the narrower range of S-option values are nested within the broader E-option values, so
403 that any given S-option has a higher expected value compared to the 4 negative E-
404 options, and a lower expected value compared to the 4 positive E-options. Such
405 configuration guarantees that participants neglecting E-option values in the ES-phase will
406 exhibit a chance-level choice accuracy (50% of expected value maximizing choices).
407 Except for the modification of the lotteries, Exp. 6 present the exact number of trials.

408 Despite this stronger economic incentive, the behavioral pattern in ES-phase remained
409 consistent with the *experiential value neglect* scenario (**Fig. 4B**). The significant
410 difference between ES and EE slopes persisted in Exp. 7 ($T(70)=5.12$, $P<0.001$),
411 suggesting that despite the reorganization of probabilities, we were still able to elicit more
412 accurate E-option values from EE-choices (**Fig.4F**, **Fig. 4G**). As a consequence,
413 compared to Exp. 6, the accuracy in the ES-choices significantly dropped in Exp 7 by
414 approximately 20% ($T(94.97)=11.01$, $P < .001$, **Fig 4C**). Of note, the accuracy in the EE-
415 choices remained the same (**Fig. 4D**, **Fig. 4E**), with no significant difference between the
416 two experiments ($T(131.77)=0.38$, $P=1$, $BF^{10}=0.19$).

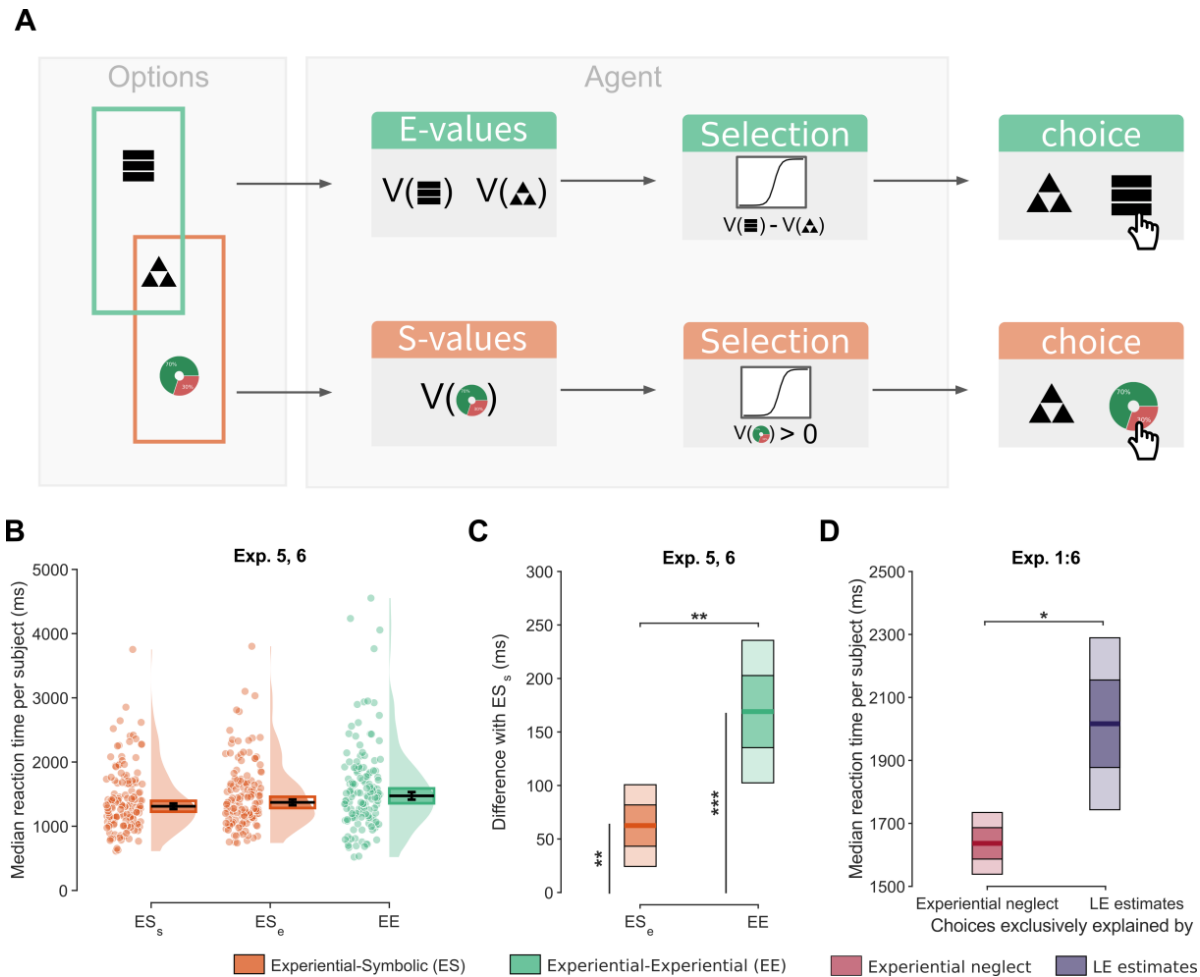
417 These findings indicate that experience values are neglected even when it involves an
418 (economic) cost. Therefore, the results are consistent with the idea that the experiential
419 value neglect reflects a hard-coded feature of hybrid choices between experiential and
420 symbolic option, rather than being strategically deployed by the relative lack of incentive
421 in Exp1-6.

422 **Controlling for ambiguity aversion**

423 E-options may be deemed more ambiguous, because their outcome probability
424 distributions are inferred from finite samples and cannot be known with absolute
425 precision or certainty. Experiential value neglect cannot be accounted by a simple form
426 of ambiguity aversion (46–48), because E-options are generally preferred compared to
427 negative expected value S-options (i.e., there is no systematic bias *against* E-options).
428 Nonetheless, to assess whether the participant's attitude toward ambiguous lotteries
429 differed between experiential and symbolic options in a final experiment we included
430 choices with ambiguous lotteries (i.e., lotteries, whose value was hidden). The results
431 (presented in the **Supplementary Materials** and **Figure S1**) indicate that ambiguity

432 aversion was not detectable in our set up and that it could therefore not contribute to
 433 explain the observed pattern of behavior. The results of Exp. 8 also replicate all previously
 434 reported findings.

435



436

437 **Fig. 5 Hypothetical decision model and reaction times analyses (A)** The panel presents a schematic
 438 representation of the decision process in the EE- and the ES- phases, respectively. The two processes
 439 differ in that in the former case (EE) the decision is based by retrieving the values of both options, while in
 440 the latter case (ES), under an extreme form of experiential value neglect, only the value of the lottery
 441 matters. **(B)** Median reaction times across modalities. EE decisions are significantly longer than ES
 442 decisions (regardless of the choice taken in ES). When comparing when an S-option is chosen (ES_s) and
 443 when an E-option is chosen (ES_e) we also observed a significant difference. The black lines represent mean
 444 and standard error of the mean. The colored boxes represent 95% confidence interval. The shaded area
 445 probability density functions. **(C)** Different in reaction times differences ($ES_e - ES_s$ in orange; $EE - ES_s$ in
 446 green). In the data-boxes, the dark tone line represents the mean, mid-dark tone the standard mean error,
 447 light tone a 95% confidence interval. **(D)** Reaction times as a function of whether the ES-choices could be
 448 only explained by a total neglect of the experiential value (red) or whether they could only be explained by
 449 experiential values estimated from the learning phase (dark blue). In the data-boxes, the dark tone
 450 line represents the mean, mid-dark tone the standard mean error, light tone a 95% confidence interval. * $p < 0.05$,
 451 ** $p < 0.01$, *** $p < 0.001$ paired two-sample t-test.

452 **Reaction times analysis: a tale of two systems?**

453 Choice behavior differ across the ES- and the EE-choices. In the ES-phase, participants
454 neglect the experiential option value and to make choices only based on the symbolic
455 option value, so that, if the S-option is positive, it is chosen, otherwise it is rejected (**Fig.**
456 **5A**). On the other hand, EE-choices are based on the retrieval from memory of the
457 experiential values of both options. Thus, one decision process (ES-choices) seems to
458 involve the processing and representation of only one option value (the lottery), while the
459 other process (EE-choices) seems to involve the processing and the representation of
460 two option values. We hypothesized that these different processes translate into different
461 reaction times between the two choice modalities. To test this prediction, we compared
462 the reaction times in EE and ES-choices, while including only decisions with similar
463 objective value difference (49). Indeed, we found that ES decisions were faster compared
464 to EE decisions, both when the S-option is chosen – (ES_s) and when the E-option is
465 chosen –(ES_e) (T(136)=6.02, P < 0.001; T(136)=3.98, P < 0.001; **Fig. 5B** and **Fig. 5C**).
466 Of note, within ES decisions, ES_e choices were also slightly but significantly slower the
467 ES_s choices (~50ms; T(136)=4.35, P < 0.001), which may indicate that choosing the E-
468 option requires additional processing to retrieve and represent the value of the E-option.
469 To confirm this intuition, we considered two categories of ES-choices: choices exclusively
470 consistent with the participant choosing using the estimates inferred from the LE phase,
471 on one side, and choices consistent with a full experiential value neglect, on the other
472 side (**Fig. S5**). We observed that, in conformity with previous results, ES-choices that are
473 consistent with a full experiential value neglect are significantly faster than choices that
474 can only be explained taking into account the E-option values estimated from the LE-
475 phase (T(386)= 2.27, P<0.05) (**Fig. 5D**). Overall, the RT analyses support the idea that
476 choices based on the S-values of the lotteries required reduced cognitive processing
477 compared to those involving the retrieving from memory. Thus, E-values inferred from
478 ES-choices are consistent with the dual process model of **Fig. 5A**.

479 **Discussion**

480 Our results clearly indicate that the experiential and symbolic option values are not treated
481 symmetrically when making hybrid choices and speak against the idea of a central

482 valuation system that encodes option values in a common currency, regardless of the
483 way they are built (3, 12). The key finding supporting this claim is provided by the analysis
484 of hybrid decision problems between experiential and symbolic cues, where choices
485 appeared to be made by largely neglecting value information acquired during the learning
486 phase. Crucially, by running several experiments and including multiple control measures,
487 we ruled out several alternative explanations for the experiential value neglect: this
488 decision-making pattern is not due to insufficient learning, forgetting, generalization issue,
489 or a lack of incentive. Finally, reaction time analyses are consistent with different
490 processing of experiential and symbolic values and with the idea of an additional cognitive
491 cost associated with the memory retrieval of learned values. It seems that past
492 experiences and symbolic descriptions of possible outcomes ultimately generate value
493 representations different enough to make them largely incommensurable and that the
494 tension between the two is resolved by overweighting (or prioritizing) symbolic
495 information. In the following paragraphs we try to provide plausible reasons why these
496 values representations radically differ, why symbolic information is favored in hybrid
497 choices and which cognitive mechanisms could underlie the behavioral pattern observed.

498 Symbolic descriptions of lotteries in our task (and in general) involve separate information
499 about at least two different features of outcomes: payoffs (i.e., the amount of reward to
500 be won or lost) and their probability (50). Models of decision-making designed to explain
501 behavior in this kind of paradigms frequently assume that probability and payoffs are
502 processed individually. For instance, in prospect theory and its extensions, different
503 subjective weighting functions are supposed to apply to these variables (51–53, 14, 54).
504 A separate representation of payoffs and probabilities is also assumed by models that do
505 not suppose the calculation of a multiplicative expected utility (55) and by models
506 supposing that decisions are underpinned by feature-by-feature comparisons (56–60).
507 On the contrary, experience-based choices, as instantiated by simple reinforcement
508 learning tasks, are usually modeled assuming that the decision-makers represents a
509 unique numeric value for each state-action pair. The decision-maker can ‘look-up’ in this
510 value matrix before making their choice and, once an outcome is obtained it partially
511 overwrites the ‘cached’ values previously stored in memory, so that they approximate the
512 average outcome (61). Option value representation is therefore structurally very different

513 from that of description-based choices, because the relevant features (payoffs and
514 probabilities) are never explicitly represented as separate attributes of the outcomes.
515 Furthermore, some authors even suggest that reinforcement-based choices may bypass
516 the calculation of reward-based option-specific values, and is underpinned by what is
517 called direct policy learning (62–65). Our results seem to reject an extremely orthodox
518 interpretation of direct policy learning (accuracy in the learning phase was sensitive to the
519 value difference between options and experiential values were successfully generalized
520 to new combinations). It is nonetheless plausible to conceive that - at least to some extent
521 - reinforcement-based decisions involve a value-free (policy-based) component that can
522 be hardly compared with the subjective extracted from explicit payoffs and probabilities.
523 Functional neuroimaging investigations of experiential and symbolic decision-making
524 may also shed light on the debate about value representation across modalities. While
525 functional meta-analyses identified overlapping correlates of experiential and symbolic
526 values (17–20), the putative neural mechanisms of reinforcement-based and description-
527 based decisions differ in many crucial respects. First of all, the most influential and
528 consensual neural models of reinforcement-based learning and decision-making give a
529 preponderant role to dopamine-induced neural plasticity circuits (66–68). More
530 specifically dopamine-dependent plasticity is supposed to drive action selection by
531 shaping the strength of the synapses between the frontal cortex and the basal ganglia
532 (69, 70). Current neural models do not attribute to dopamine-driven processes and the
533 basal ganglia a prominent role in description-based choices. Rather, they suppose that
534 the decision process is solved by cortical circuits (71–74), following an evidence
535 accumulation process similar to that observed for perceptual decisions (75, 76). Thus,
536 structural differences in the neural mechanisms of choices across modalities may
537 represent a biologically grounded bases of the representational difference between
538 experiential and symbolic values.

539 The representational tension of hybrid choices is solved by participants by neglecting the
540 experiential values and basing their choices on the symbolic value. Several control
541 analyses allowed us to formally exclude the possibility that this effect merely arise from
542 insufficient knowledge of the experiential values. Why is the symbolic information
543 preferred? We suggest two not-mutually exclusive explanations. One possibility is that

544 experiential value estimates are perceived as less precise. Note here that precision
545 represents the uncertainty about the value estimate itself (48). Indeed, assuming
546 imperfect memory storage and retrieval, it is conceivable that experiential values are less
547 precise compared to symbolic ones that can be perfectly calculated (77). According to
548 this interpretation, participants would quasi-systemically prioritize the more precise
549 source of information for their choices (47, 48, 78). Another possibility is that participants
550 prefer discarding experiential information not to incur the cost associated with the cost of
551 memory retrieval (79, 80). Reaction times analysis was overall consistent with this idea,
552 because choices involving the processing of the experiential values were generally slower
553 compared to those involving symbolic ones, even if balanced in objective difficulty (49).
554 This latter interpretation leaves open the possibility that if one makes memory retrieval
555 less costly, the behavioral pattern could be reversed (i.e., we would witness symbolic
556 value neglect). This could be possible for example after extensive training, once
557 experience-based choices are routinized (81) or, conversely, by making symbolic
558 information harder to decode. These are interesting possibilities to be explored by future
559 studies.

560 Finally, we speculate on the possible cognitive mechanisms underlying the experiential
561 value neglect phenomenon and we identify two plausible candidates. The first mechanism
562 involves 'bottom-up' attentional processes. It is well-documented that attentional focus
563 biases evidence accumulation in value based decision-making (82, 83). It is therefore
564 conceivable that an attentional bias toward symbolic options may result in prioritizing
565 described information and neglecting experiential one. The second possible mechanism
566 involves a 'top-down' heuristic process, according to which the calculation of individual
567 option values is hijacked by a deterministic decision rules (44). Of note, even if we
568 managed to demonstrate experiential value neglect in situations where it is
569 disadvantageous (experiment 7), it can nonetheless be argued that this decision rule is
570 overall adaptive, because computationally cheap and satisfying in most situations (see
571 experiments 1-6).

572 To conclude, our results add to the collection of behavioral anomalies showing that
573 values representations are inherently dependent on the way they are built, as it is

574 postulated by the ‘construction of preference’ framework (84, 14, 85). More specifically,
575 our findings pose serious challenges to the default assumption that values
576 representations are shared across different decision-making modalities, traditionally
577 referred to as experience- and description-based. The incommensurability between
578 experiential and symbolic values results in behaving as if discarding acquired information
579 and consequently entails suboptimal decisions. These findings are worth exploring
580 outside the experimental setting because many real-life decisions involve a tension
581 between an experiential and a symbolic component.

582

583 **Acknowledgements**

584 The authors thank Aurélien Baillon for helpful comments. SP is supported by the Institut
585 de Recherche en Santé Publique (IRESF, grant number : 2011138-00), the Agence
586 National de la Recherche (CogFinAgent: ANR-21-CE23-0002-02; RELATIVE: ANR-21-
587 CE37-0008-01; RANGE : ANR-21-CE28-0024-01) and the Ville de Paris (Emergence(s)).
588 ML is supported by an SNSF Ambizione grant (PZ00P3_174127) and an ERC Starting
589 Grant (948671). The article was prepared in the framework of a research grant funded by
590 the International Laboratory for Social Neuroscience of the Institute for Cognitive
591 Neuroscience HSE (RF Government grant No 075-15-2019-1930) and by the
592 Departement d'études cognitives (FrontCog ANR-17-EURE-0017).

593

594 **References**

- 595 1. P. A. Samuelson, A Note on the Pure Theory of Consumer's Behaviour. *Economica*. **5**, 61–
596 71 (1938).
- 597 2. J. Von Neumann, O. Morgenstern, *Theory of games and economic behavior* (Princeton
598 University Press, Princeton, NJ, US, 1944), *Theory of games and economic behavior*.
- 599 3. A. Rangel, C. Camerer, P. R. Montague, A framework for studying the neurobiology of
600 value-based decision making. *Nat. Rev. Neurosci.* **9**, 545–556 (2008).
- 601 4. R. J. Herrnstein, Relative and absolute strength of response as a function of frequency of
602 reinforcement. *J. Exp. Anal. Behav.* **4**, 267 (1961).
- 603 5. B. F. Skinner, *Science and human behavior* (Simon and Schuster, 1965).
- 604 6. R. S. Sutton, A. G. Barto, *Reinforcement learning: An introduction* (MIT press, 2018).
- 605 7. D. Bernoulli, Exposition of a New Theory on the Measurement of Risk. *Econometrica*. **22**,
606 23–36 (1738).
- 607 8. P.-S. Laplace, *Essai philosophique sur les probabilités* (H. Remy, 1829).
- 608 9. P. Wakker, A. Tversky, An axiomatization of cumulative prospect theory. *J. Risk Uncertain.*
609 **7**, 147–175 (1993).
- 610 10. D. Kahneman, A. Tversky, in *Handbook of the fundamentals of financial decision making:*
611 *Part I* (World Scientific, 2013), pp. 269–278.
- 612 11. B. De Martino, D. Kumaran, B. Seymour, R. J. Dolan, Frames, biases, and rational
613 decision-making in the human brain. *Science*. **313**, 684–687 (2006).
- 614 12. P. W. Glimcher, *Foundations of neuroeconomic analysis* (OUP USA, 2011).
- 615 13. C. F. Camerer, A review essay about Foundations of Neuroeconomic Analysis by Paul
616 Glimcher. *J. Econ. Lit.* **51**, 1155–82 (2013).
- 617 14. I. Vlaev, N. Chater, N. Stewart, G. D. A. Brown, Does the brain calculate value? *Trends*
618 *Cogn. Sci.* **15**, 546–554 (2011).
- 619 15. N. Stewart, EPS Prize Lecture: Decision by sampling: The role of the decision environment
620 in risky choice. *Q. J. Exp. Psychol.* **62**, 1041–1062 (2009).
- 621 16. I. Erev, E. Ert, O. Plonsky, D. Cohen, O. Cohen, From anomalies to forecasts: Toward a
622 descriptive model of decisions under risk, under ambiguity, and from experience. *Psychol.*
623 *Rev.* **124**, 369 (2017).
- 624 17. O. Bartra, J. T. McGuire, J. W. Kable, The valuation system: a coordinate-based meta-
625 analysis of BOLD fMRI experiments examining neural correlates of subjective value.
626 *Neuroimage*. **76**, 412–427 (2013).

- 627 18. J. Garrison, B. Erdeniz, J. Done, Prediction error in reinforcement learning: a meta-
628 analysis of neuroimaging studies. *Neurosci. Biobehav. Rev.* **37**, 1297–1310 (2013).
- 629 19. J. A. Clithero, A. Rangel, Informatic parcellation of the network involved in the computation
630 of subjective value. *Soc. Cogn. Affect. Neurosci.* **9**, 1289–1302 (2014).
- 631 20. E. Fouragnan, C. Retzler, M. G. Philiastides, Separate neural representations of prediction
632 error valence and surprise: Evidence from an fMRI meta-analysis. *Hum. Brain Mapp.* **39**,
633 2887–2906 (2018).
- 634 21. R. Hertwig, I. Erev, The description–experience gap in risky choice. *Trends Cogn. Sci.* **13**,
635 517–523 (2009).
- 636 22. C. R. Madan, E. A. Ludvig, M. L. Spetch, The role of memory in distinguishing risky
637 decisions from experience and description. *Q. J. Exp. Psychol.* **70**, 2048–2059 (2017).
- 638 23. D. U. Wulff, M. Mergenthaler-Canseco, R. Hertwig, A meta-analytic review of two modes of
639 learning and the description-experience gap. *Psychol. Bull.* **144**, 140 (2018).
- 640 24. B. Garcia, F. Cerrotti, S. Palminteri, The description–experience gap: a challenge for the
641 neuroeconomics of decision-making under uncertainty. *Philos. Trans. R. Soc. B.* **376**,
642 20190665 (2021).
- 643 25. D. Kellen, T. Pachur, R. Hertwig, How (in)variant are subjective representations of
644 described and experienced risk and rewards? *Cognition.* **157**, 126–138 (2016).
- 645 26. I. Erev, E. Ert, A. E. Roth, E. Haruvy, S. M. Herzog, R. Hau, R. Hertwig, T. Stewart, R.
646 West, C. Lebiere, A choice prediction competition: Choices from experience and from
647 description. *J. Behav. Decis. Mak.* **23**, 15–47 (2010).
- 648 27. T. H. B. FitzGerald, B. Seymour, D. R. Bach, R. J. Dolan, Differentiable Neural Substrates
649 for Learned and Described Value and Risk. *Curr. Biol.* **20**, 1823–1829 (2010).
- 650 28. S. R. Heilbronner, B. Y. Hayden, The description-experience gap in risky choice in
651 nonhuman primates. *Psychon. Bull. Rev.* **23**, 593–600 (2016).
- 652 29. W. M. DuCharme, M. L. Donnell, Intrasubject comparison of four response modes for
653 “subjective probability” assessment. *Organ. Behav. Hum. Perform.* **10**, 108–117 (1973).
- 654 30. R. A. Rescorla, A theory of Pavlovian conditioning: Variations in the effectiveness of
655 reinforcement and nonreinforcement. *Curr. Res. Theory*, 64–99 (1972).
- 656 31. T. E. J. Behrens, M. W. Woolrich, M. E. Walton, M. F. S. Rushworth, Learning the value of
657 information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
- 658 32. A. G. E. Collins, The Tortoise and the Hare: Interactions between Reinforcement Learning
659 and Working Memory. *J. Cogn. Neurosci.* **30**, 1422–1432 (2018).
- 660 33. S. Palminteri, M. Khamassi, M. Joffily, G. Coricelli, Contextual modulation of value signals
661 in reward and punishment learning. *Nat. Commun.* **6**, 1–14 (2015).

- 662 34. S. Bavard, A. Rustichini, S. Palminteri, Two sides of the same coin: Beneficial and
663 detrimental consequences of range adaptation in human reinforcement learning. *Sci. Adv.*
664 **7**, eabe0340.
- 665 35. G. M. Becker, M. H. DeGroot, J. Marschak, Measuring utility by a single-response
666 sequential method. *Behav. Sci.* **9**, 226–232 (1964).
- 667 36. J. Rieskamp, P. E. Otto, SSL: A Theory of How People Learn to Select Strategies. *J. Exp.*
668 *Psychol. Gen.* **135**, 207–236 (2006).
- 669 37. L. J. Savage, *The foundations of statistics* (Courier Corporation, 1972).
- 670 38. B. L. Lipman, Information Processing and Bounded Rationality: A Survey. *Can. J. Econ.*
671 *Rev. Can. Econ.* **28**, 42–67 (1995).
- 672 39. V. M. Chase, R. Hertwig, G. Gigerenzer, Visions of rationality. *Trends Cogn. Sci.* **2**, 206–
673 214 (1998).
- 674 40. G. Gigerenzer, W. Gaissmaier, Heuristic decision making. *Annu. Rev. Psychol.* **62**, 451–
675 482 (2011).
- 676 41. B. Mackowiak, F. Matejka, M. Wiederholt, Rational inattention: A review (2021).
- 677 42. H. A. Simon, Theories of bounded rationality. *Decis. Organ.* **1**, 161–176 (1972).
- 678 43. H. A. Simon, A. Newell, Human problem solving: The state of the theory in 1970. *Am.*
679 *Psychol.* **26**, 145 (1971).
- 680 44. G. E. Gigerenzer, R. E. Hertwig, T. E. Pachur, *Heuristics: The foundations of adaptive*
681 *behavior*. (Oxford University Press, 2011).
- 682 45. H. A. Simon, *Administrative behavior* (Simon and Schuster, 2013).
- 683 46. D. Ellsberg, Risk, Ambiguity, and the Savage Axioms. *Q. J. Econ.* **75**, 643–669 (1961).
- 684 47. D. Frisch, J. Baron, Ambiguity and rationality. *J. Behav. Decis. Mak.* **1**, 149–157 (1988).
- 685 48. C. Camerer, M. Weber, Recent developments in modeling preferences: Uncertainty and
686 ambiguity. *J. Risk Uncertain.* **5**, 325–370 (1992).
- 687 49. I. Krajbich, B. Bartling, T. Hare, E. Fehr, Rethinking fast and slow based on a critique of
688 reaction-time reverse inference. *Nat. Commun.* **6**, 7455 (2015).
- 689 50. C. A. Holt, S. K. Laury, Risk aversion and incentive effects. *Am. Econ. Rev.* **92**, 1644–1655
690 (2002).
- 691 51. A. Tversky, D. Kahneman, Prospect theory: An analysis of decision under risk.
692 *Econometrica.* **47**, 263–291 (1979).
- 693 52. A. Tversky, D. Kahneman, Advances in prospect theory: Cumulative representation of
694 uncertainty. *J. Risk Uncertain.* **5**, 297–323 (1992).

- 695 53. D. Prelec, The Probability Weighting Function. *Econometrica*. **66**, 497 (1998).
- 696 54. J. Quiggin, *Generalized expected utility theory: The rank-dependent model* (Springer
697 Science & Business Media, 2012).
- 698 55. A. Soltani, E. Koechlin, Computational models of adaptive behavior and prefrontal cortex.
699 *Neuropsychopharmacology*, 1–14 (2021).
- 700 56. K. Rayner, Eye movements in reading and information processing: 20 years of research.
701 *Psychol. Bull.* **124**, 372 (1998).
- 702 57. A. Glöckner, A.-K. Herbold, An eye-tracking study on information processing in risky
703 decisions: Evidence for compensatory strategies based on automatic processes. *J. Behav.*
704 *Decis. Mak.* **24**, 71–98 (2011).
- 705 58. S. Fiedler, A. Glöckner, The Dynamics of Decision Making in Risky Choice: An Eye-
706 Tracking Analysis. *Front. Psychol.* **3**, 335 (2012).
- 707 59. V. Venkatraman, J. W. Payne, S. A. Huettel, An overall probability of winning heuristic for
708 complex risky decisions: Choice and eye fixation evidence. *Organ. Behav. Hum. Decis.*
709 *Process.* **125**, 73–87 (2014).
- 710 60. J. A. Aimone, S. Ball, B. King-Casas, It's not what you see but how you see it: Using eye-
711 tracking to study the risky decision-making process. *J. Neurosci. Psychol. Econ.* **9**, 137–
712 144 (2016).
- 713 61. R. S. Sutton, Learning to predict by the methods of temporal differences. *Mach. Learn.* **3**,
714 9–44 (1988).
- 715 62. P. Dayan, L. F. Abbott, *Theoretical neuroscience: computational and mathematical*
716 *modeling of neural systems* (Computational Neuroscience Series, 2001).
- 717 63. J. Li, N. D. Daw, Signals in Human Striatum Are Appropriate for Policy Update Rather than
718 Value Prediction. *J. Neurosci.* **31**, 5504–5511 (2011).
- 719 64. B. Y. Hayden, Y. Niv, The case against economic values in the orbitofrontal cortex (or
720 anywhere else in the brain). *Behav. Neurosci.* **135**, 192 (2021).
- 721 65. D. Bennett, Y. Niv, A. J. Langdon, Value-free reinforcement learning: policy optimization as
722 a minimal model of operant behavior. *Curr. Opin. Behav. Sci.* **41**, 114–121 (2021).
- 723 66. M. J. Frank, C. D'Lauro, T. Curran, Cross-task individual differences in error processing:
724 neural, electrophysiological, and genetic components. *Cogn. Affect. Behav. Neurosci.* **7**,
725 297–308 (2007).
- 726 67. A. G. Collins, M. J. Frank, Opponent actor learning (OpAL): modeling interactive effects of
727 striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev.* **121**, 337
728 (2014).
- 729 68. M. Möller, R. Bogacz, Learning the payoffs and costs of actions. *PLOS Comput. Biol.* **15**,
730 e1006285 (2019).

- 731 69. P. Redgrave, T. J. Prescott, K. Gurney, The basal ganglia: a vertebrate solution to the
732 selection problem? *Neuroscience*. **89**, 1009–1023 (1999).
- 733 70. I. Bar-Gad, H. Bergman, Stepping out of the box: information processing in the neural
734 networks of the basal ganglia. *Curr. Opin. Neurobiol.* **11**, 689–695 (2001).
- 735 71. L. T. Hunt, N. Kolling, A. Soltani, M. W. Woolrich, M. F. Rushworth, T. E. Behrens,
736 Mechanisms underlying cortical activity during value-guided choice. *Nat. Neurosci.* **15**,
737 470–476 (2012).
- 738 72. A. Rustichini, C. Padoa-Schioppa, A neuro-computational model of economic decisions. *J.*
739 *Neurophysiol.* **114**, 1382–1398 (2015).
- 740 73. C. Padoa-Schioppa, K. E. Conen, Orbitofrontal cortex: a neural circuit for economic
741 decisions. *Neuron*. **96**, 736–754 (2017).
- 742 74. S. Farashahi, C. H. Donahue, P. Khorsand, H. Seo, D. Lee, A. Soltani, Metaplasticity as a
743 Neural Substrate for Adaptive Learning and Choice under Uncertainty. *Neuron*. **94**, 401-
744 414.e6 (2017).
- 745 75. A. C. Huk, M. N. Shadlen, Neural Activity in Macaque Parietal Cortex Reflects Temporal
746 Integration of Visual Motion Signals during Perceptual Decision Making. *J. Neurosci.* **25**,
747 10420–10436 (2005).
- 748 76. C. A. Hutcherson, B. Bushong, A. Rangel, A neurocomputational model of altruistic choice
749 and its implications. *Neuron*. **87**, 451–462 (2015).
- 750 77. C. Findling, N. Chopin, E. Koechlin, Imprecise neural computations as a source of adaptive
751 behaviour in volatile environments. *Nat. Hum. Behav.* **5**, 99–112 (2021).
- 752 78. R. Bricet, “Preferences for information precision under ambiguity” (THEMA (THéorie
753 Economique, Modélisation et Applications), Université de ..., 2018).
- 754 79. R. Dukas, Costs of Memory: Ideas and Predictions. *J. Theor. Biol.* **197**, 41–50 (1999).
- 755 80. H. Afrouzi, S. Kwon, Y. Ma, “A Model of Costly Recall” (working paper, Columbia
756 University, 2020).
- 757 81. K. J. Miller, A. Shenhav, E. A. Ludvig, Habits without values. *Psychol. Rev.* **126**, 292
758 (2019).
- 759 82. I. Krajbich, C. Armel, A. Rangel, Visual fixations and the computation and comparison of
760 value in simple choice. *Nat. Neurosci.* **13**, 1292–1298 (2010).
- 761 83. P. Sepulveda, M. Usher, N. Davies, A. A. Benson, P. Ortoleva, B. De Martino, Visual
762 attention modulates the integration of goal-relevant evidence and not value. *eLife*. **9**,
763 e60705 (2020).
- 764 84. S. Lichtenstein, P. Slovic, *The construction of preference* (Cambridge University Press,
765 2006).

- 766 85. B. Hayden, Y. Niv, The case against economic values in the brain (2020),
767 doi:10.31234/osf.io/7hgup.
- 768 86. R. D. Luce, The choice axiom after twenty years. *J. Math. Psychol.* **15**, 215–233 (1977).
- 769 87. R. D. Luce, *Individual choice behavior: A theoretical analysis* (Courier Corporation, 2012).
- 770 88. M. Lebreton, K. Bacily, S. Palminteri, J. B. Engelmann, Contextual influence on confidence
771 judgments in human reinforcement learning. *PLoS Comput. Biol.* **15**, e1006973 (2019).
- 772 89. R. C. Wilson, A. G. Collins, Ten simple rules for the computational modeling of behavioral
773 data. *eLife.* **8**, e49547 (2019).

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796 **Methods and supplementary results**

797 In this document we present the methods, as well as some additional results, including
798 those issued from an experiment (Exp. 8), which is only briefly mentioned in the main text.

799 **Experimental participants**

800 In total, we tested 787 participants (430 females; aged 31.09 ± 10.42 years) distributed
801 across seven experiments. Participants were recruited via Prolific, a platform dedicated
802 to online research participants recruitment (<https://prolific.co/>). To assess participants'
803 engagement in the different tasks and their understanding of probability representation,
804 we inserted catch trials consisting in choices between two lotteries (S-options), with one
805 of the two cues being obviously better in terms of expected value maximization. In all
806 analyses we only retained the participants displaying 100% of correct choices in these
807 catch trials. In total 599 participants were included. Experiment 1 to 7 included the
808 following numbers of participants: 76, 71, 83, 88, 71, 66, 71, 73 (see **Table 1**). Of note,
809 none of the results presented in the main or supplemental text was affected by the
810 exclusion of the participants.

| Exp. | Outcome (LE) | Structure (LE) | Decision problems (LE) | Phases | Sessions | N |
|------|--------------|----------------|------------------------|-------------------|----------|----|
| 1 | partial | interleaved | 4 | LE ES SP | 1 | 76 |
| 2 | partial | blocked | 4 | LE ES SP | 1 | 71 |
| 3 | complete | blocked | 4 | LE ES SP | 1 | 83 |
| 4 | complete | blocked | 2 | LE ES SP | 1 | 88 |
| 5 | complete | blocked | 4 | LE ES EE SP | 1 | 71 |
| 6 | complete | blocked | 4 | LE ES EE SP | 2 | 66 |
| 7 | complete | blocked | 4 | LE ES EE SP | 2 | 71 |
| 8 | complete | blocked | 4 | LE ES EE SP EA/SA | 2 | 73 |

811

812 **Table S1. Experiments parameters.** The 'Exp.' column refers to the experiment number. The 'Outcome (LE)' column refers to the
813 outcomes displayed during a single LE phase trial. The column can take two values: partial (only obtained outcome) or complete (both
814 obtained and forgone outcomes). The 'Structure (LE)' column refers to how the presentation of the options (or decision problems) was
815 organized in the LE phase. 'Blocked' correspond to the case in which all trials belonging to a given option pair are presented in a row.
816 Otherwise, when options pairs are distributed randomly, the value is set to 'interleaved'. The 'Decision problems (LE)' column refers
817 to the number of option pairs presented in the LE phase. The 'Phases' column, refers to the specific phases present in a particular
818 experiment. 'LE' refers to the learning phase. 'ES' stands for Experiential-Symbolic phase. 'EE' stands for Experiential-Experiential
819 phase (performed after learning with no feedback). 'SP' stands for Stated Probability phase. 'EA/SA' stands for Experiential-
820 Ambiguous and Symbolic-Ambiguous. The 'Sessions' column provides the number of sessions, i.e., how many times we repeated
821 the sequence of phases with a different set of E-options. The 'N' column refers to the number of participants included in the experiment
822 after exclusion of those displaying >100% correct response rate in the ES catch trials.

823 The research was carried out following the principles and guidelines for experiments
824 including human participants provided in the declaration of Helsinki (1964, revised in
825 2013). The INSERM Ethical Committee approved the study and participants provided
826 written informed consent prior to their inclusion. To sustain motivation throughout the
827 experiment, the tasks were economically incentivized. Specifically, in addition to a show-
828 up fee, participants were initially endowed with £2.5, and according to their choices, they
829 could reach a maximum £5. The conversation rate was around 1pt = 1 cent and they were
830 explained that all points won across the different phases were summed up. The average
831 final bonus was £4.05 ± 0.72, which was significantly higher compared to what they would
832 have got in average following random choices ($T(615) = 52.58, P < 0.001$).

833 Behavioral protocol

834 The different experiments were conducted on a website programmed in javascript, html
835 and css (code: <https://github.com/bsgarcia/RetrieveAndCompare>, testing: <https://human-rl.scicog.fr/RandCTesting>).

837

838 **Initial learning phase (LE phase)**

839 Participants first performed a probabilistic instrumental learning task (LE). Participants
840 were provided with written instructions explaining that the aim of the task was to maximize
841 their payoff by seeking monetary rewards and avoiding monetary losses. From
842 experiment 1 to 5, participants performed only one learning session. Experiment 6, 7 and
843 8 for their part include 2 learning sessions. From experiment 1 to 7, each learning session
844 contained four pairs of experiential cues (E-options), apart from experiment 4, which
845 contained 2 (but featured proportionally twice more trials). Each pair was fixed, so that a
846 given cue was always presented against the same other cue. Thus, within learning
847 sessions, pairs of cues represented stable choice contexts. Within each pair, the two cues
848 were associated to two outcomes; either winning a point (+1) either losing one (-1). The
849 four (two in experiment 4) cue pairs corresponded to four contexts of varying difficulty,
850 indexed by the difference in the probability of winning a point between the two cues. On
851 each trial, one pair was randomly presented with one cue on the right and the other on
852 left side of the screen. Participants were required to select, without time-limit, between
853 the two cues by left-clicking. After the choice, the selected cue was highlighted with a
854 black border while a transition effect was activated. The transition effect lasted
855 approximately 1000 ms and revealed the outcome of the choice. The outcome was then
856 displayed during approximately 1500 ms. In experiments 1, 2, 3, 5, 6 and 7, the four pair
857 of cues were presented 30 times each, for a total of 120 trials within sessions. In
858 experiment 4, the two pairs were presented 60 times each, to maintain an identical
859 number of trials. In experiment 1, pairs of cues were presented in an interleaved manner,
860 meaning they were distributed randomly across the 120 trials. From experiment 2 to 7,
861 pairs were presented in a blocked manner, meaning they were stacked in sequences of
862 30 choices.

863 Regarding feedback, there were two settings: partial and complete. A partial feedback
864 setting implied that only the outcome of the chosen option (or cue) was displayed, while
865 complete feedback means that both outcomes were displayed, regardless of the choice.
866 Experiment 1 and 2 involved partial feedback. From experiment 3 on, feedback was set
867 to complete.

868 **Hybrid choices between experiential and symbolic values (ES phase)**

869 This phase is present in all the experiments.

870 After the LE phase, E-options were presented against symbolic cues (S-options). S-
871 options were implemented as pie-charts, where the green part indicates the probability to
872 win a point, and the red part indicates the probability to lose a point. Each E-option (8)
873 involved in the LE phase was presented against 11 S-options (for a total of 88 trials), with
874 probability of winning (and respectively loosing) a point ranging from 0% to 100%, with a
875 10% step. On each trial, one pair was randomly presented with one cue on right and left
876 side of the screen. Participants were required to select, without time-limit, between the
877 two cues by left-clicking. After the choice, the selected cue was highlighted with a black
878 border and the transition to the next trial, lasted approximately 1000 ms. No feedback
879 was presented during the ES phase. Participants were informed about their earnings only
880 at the end.

881 Although the outcome was not displayed, participants were told that this phase was still
882 incentivized, such that choice accuracy affected their bonus compensation.

883 **Assessing generalization of experiential values (EE phase)**

884 This phase is present in Experiment 5 to 8. After the LE phase, each E-option was
885 presented against other E-options. With 8 cues presented in the LE phase, it follows that
886 each E-option was presented against the other 7 E-options, so that this phase contained
887 56 trials. EE choices were presented in the same time as the ES choices, because we
888 wanted to avoid having them differ in terms of time elapsed since the LE phase. Thus,
889 technically the EE and the ES phases overlap.

890 For each trial, one pair was randomly presented with one cue on the right and left sides
891 of the screen. Participants were required to select, without time-limit, between the two
892 cues by left-clicking. After the choice, the selected cue was highlighted with a black border
893 and the transition to the next trial, lasted approximately 1000 ms. The transition effect
894 lasted approximately 1000 ms and leave place for the next trial. No feedback was
895 presented.

896 Although the outcome was not displayed, participants were told that they could still win
897 (and lose) points during this phase, this phase was still incentivized, such that choice
898 accuracy affected their bonus compensation.

899

900

901

902 **Stated Probability assessment (SP phase)**

903 In all experiments, participants were asked, for each E-option previously faced in the LE
904 phase, the following question «What are the odds this symbol gives a +1?». They had to
905 provide their answer on rating-scale, going from 0% to 100% with a 5% step.

906 Answers were incentivized via a matching probability procedure that is based on the
907 Matching Probability Mechanism (29). More precisely, participant chose a probability (p)
908 for the presented E-option. A number (r) is then randomly drawn in the interval $[0, 1]$. If p
909 $> r$, the outcome of the choice was obtained using the E-option probability of winning and
910 losing a point (as-if the E-option was chosen in the LE phase for instance). Otherwise, if
911 $p < r$, the participant has r (%) chance of winning a point, and respectively $1-r$ (%) chance
912 of losing a point.

913 In other words, the higher the response (p) of the participant, the higher the chances were
914 the outcome would be determined by the E-option. Conversely, the lower the response
915 (p), the higher the chances were that the outcome would be determined by the random
916 lottery number (r).

917 **Ambiguity assessment**

918 Preference towards ambiguous lotteries was assessed only in Experiment 8. After the LE
919 phase, E-options as well as S-options were presented against an ambiguous cue. This
920 ambiguous cue was represented by a greyed pie-chart, with a question mark on top.
921 Consequently, it was represented similarly to S-options, i.e., as a lottery, which was
922 however 100% ambiguous in the sense that it conveys no a priori information regarding
923 probabilities of gains or losses (see **Figure S1**). Each E-option (8) and S-options (8),
924 were presented against this ambiguous cue two times, resulting in a total of 32 trials. For
925 each trial, one pair was randomly presented with one cue on the right and left sides of the
926 screen. Participants were required to select, without time-limit, between the two cues by
927 left-clicking. After the choice, the selected cue was highlighted with a black border while
928 a transition effect was activated. The transition effect lasted approximately 1000 ms and
929 left room for the next trial. No feedback was presented.

930 Although the outcome was not displayed, participants were told that they still could win
931 (and lose) points during this phase, and that correct choices (i.e., choices maximizing
932 expected value) and wrong choices would thus affect their bonus compensation.

933 **Statistical and computational modeling**

934 **Inferential statistics**

935 All t-tests were realized using Python 3.9 and the *pairwise_ttests* function from the
936 pingouin library. Bonferroni's corrections were applied systematically. Linear regressions
937 were realized using Matlab R2020a *fitlm* function.

938 **E-option probabilities inference in ES and EE phases**

939 To infer a probability estimate (or indifference point) for each E-option from EE and ES
940 choices we proceeded as follows. In those phases, an E-option was assessed relatively
941 to other cues (either S-options, either other E-options). In the ES phase 11 S-options were
942 presented against each E-option. In the EE phase 7 E-options were presented against
943 each E-option. Choosing the E-option that was currently assessed is always coded as 1,

944 whereas choosing the cue presented against (an S-option in ES, or an E-option in EE) is
945 always coded as 0.

946 We note $c_i^t \in \{0, 1\}$ the choice of a participant i at trial t . Thus, for each E-option j we
947 obtain a vector of choices $C_i^j = (c_i^1, c_i^2, c_i^3, \dots, c_i^n)$, with $n = 11$ in the ES phase, and $n = 7$
948 in the EE phase. We then fit the following logistic function (86, 87):

$$949 \quad f(C_i^j) = \frac{1}{1 + e^{\beta_i(C_i^j - \lambda_i^j)}}$$

950 With $\beta_i > 0$ (which controls the slope of the function) being a free parameter unique to
951 each individual i , while $\lambda_i^j \in [0, 1]$ (the function midpoint) is a free parameter that is
952 estimated for each E-option j and individual i . The indifference point λ_i^j represents here
953 the probability where a preference shift (from one cue to another) occurs, and is thus a
954 subjective probability (or value) estimate for the E-option j and participant i . Both
955 parameters were estimated through minimum negative log-likelihood estimation, using
956 matlab's *fmincon* function.

957 **Inferring E-option value estimated in the LE phase**

958 To infer E-option values in the learning (LE) phase, we fitted a reinforcement learning
959 model (or Q-learning model) to our data (31, 77).

960 The model treats each pair of cues as a state s . After a choice, each cue subjective
961 probability of winning a point (p_{win}) was incrementally updated with the following
962 Rescorla-Wagner rule:

$$963 \quad p_{win}(s, c) \leftarrow p_{win}(s, c) + \alpha \delta_c$$

$$964 \quad p_{win}(s, u) \leftarrow p_{win}(s, u) + \alpha \delta_u$$

965 Where α is the learning rate (which controls to what extent new information overrides
966 previous one) for the chosen cue (c) as well as the unchosen cue (u). The associated
967 prediction errors δ_c and δ_u are computed as follows:

$$968 \quad \delta_c = R_c - p_{win}(s, c)$$

969
$$\delta_u = R_u - p_{win}(s, u)$$

970 Where R_c and R_u are the outcomes displayed for both chosen and unchosen cues. R_x
 971 took value of 1, when the outcome was +1pt, and 0 otherwise. Initial were set at 0.5 for
 972 all options. Please note that for experiments 1 and 2, where only R_c was displayed (partial
 973 feedback setting) only $p_{win}(s, c)$ was updated. Decision was modeled using a softmax
 974 function, where the actual probability of choosing a cue a when presented against a cue
 975 b was calculated as follows:

976
$$P(s, a) = \frac{1}{1 + e^{\beta(p_{win}(s,b) - p_{win}(s,a))}}$$

977 With $\beta > 0$ being the temperature parameter, that implements choice stochasticity. As β
 978 decreases, the events of choosing a or b tend to become equi-probable. As β increases,
 979 the difference between $p_{win}(s, a)$ and $p_{win}(s, b)$ is amplified, and the choice becomes
 980 more and more deterministic (until the function almost acts as an argmax policy).

981 **Model fitting**

982 Learning rate and temperature parameters (here denoted θ) involved in the reinforcement
 983 learning model were estimated by finding values that minimized the negative logarithm of
 984 the posterior probability over the free parameters ($-\log(P(\theta|D))$), which was computed
 985 as follows:

986
$$-\log(P(\theta|D)) \propto -\log(P(D|\theta)) - \log(P(\theta))$$

987 Where $P(D|\theta)$ is likelihood of the data (i.e., the observed choices during the LE phase)
 988 given certain parameter values, and $P(\theta)$ is the prior probability of those parameter
 989 values.

990 The prior probability distribution over the learning rates was assumed as beta distributed
 991 and quasi-uniform (betapdf(1.1, 1.1)). The softmax temperature was for its part assumed
 992 to be gamma distributed (gampdf(1.2, 5)).

993 The optimization procedure was again performed using Matlab's *fmincon* function and
 994 previously described in Lebreton et al., 2019.

995 Parameter and choice recovery in EE and ES phases

996 To quantify and statistically compare the differences in preferences observed in the ES
997 and EE phase, we estimated for each E-option its theoretical subjective value (expressed
998 in terms of probability of winning a point). This value is itself inferred in term of indifference
999 points. For instance, in ES choices, one E-option with 80% chance of winning will be
1000 compared to range of S-options (going from 0% to 100% chance of winning a point). The
1001 indifference point for the E-option considered is then the S-option value at which a
1002 preference shift occurs between the two kinds of options (let's say, when the S-option is
1003 above 80%, considering the decision-maker is rational). To infer those indifference points,
1004 we fitted a logistic function to each subject choice history for each E-options in both EE
1005 and ES phases (see the methods section). We treated these indifference points as
1006 proxies for subjective values, i.e., E-option value estimates (or probability estimates, as
1007 in the numerical space considered they are equivalent).

1008 To assert that this fitting procedure is robust, and that we do not elicit random subjective
1009 values, we followed a parameter recovery procedure (89).

1010 We simulated EE and ES choices based on the (EE and ES) E-option value estimated
1011 from experiment 1 to 6.

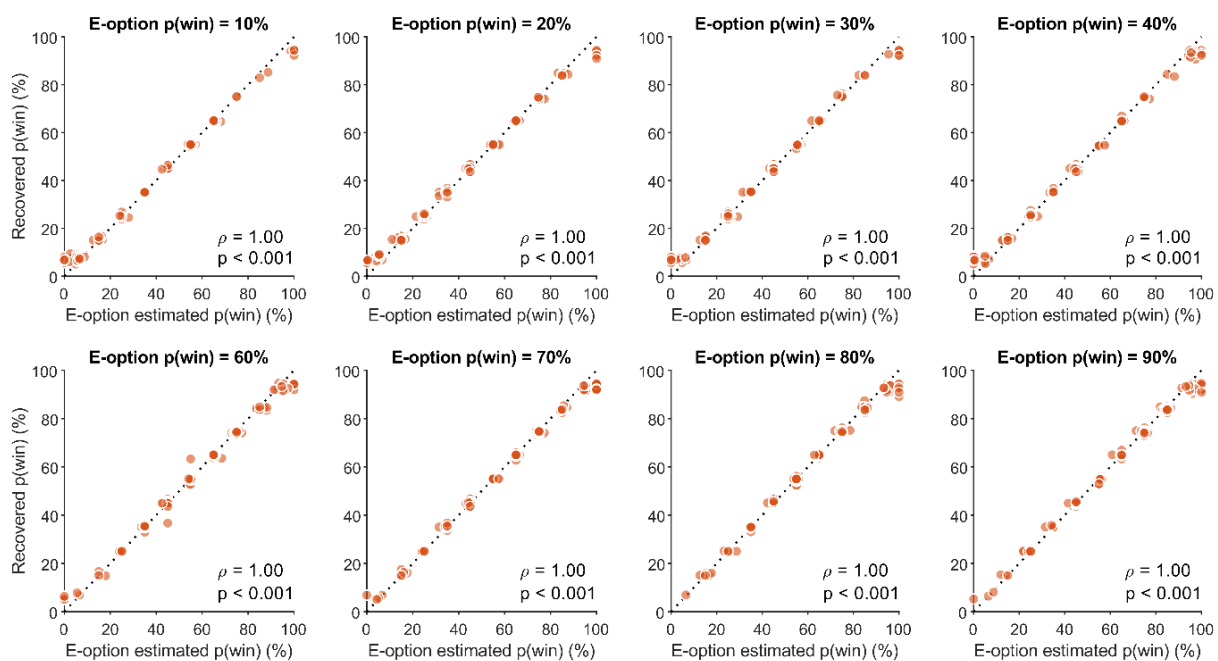
1012 More precisely, for each subject, we simulated an agent going through its choice history,
1013 and we used the 8 inferred estimates (one for each E-option, each subject having its own
1014 8 indifferent points) to simulate new choices.

1015 We generate these choices using an *argmax decision rule*, meaning that the agent
1016 systematically selects the option with the highest value. Of note, in the simulated ES
1017 phase, we do not suppose any subjective deformation regarding S-options, such that the
1018 agent is directly informed of the objective expected-value to make its decision.

1019 At this point of the procedure, simulated choices can be compared to choices from
1020 behavioral data. By doing so, we can see how well they match, and therefore whether
1021 our value estimates allow us to correctly recover the choices actually made by our
1022 participants. EE choices are recovered up to 83%, whereas ES choices are almost

1023 perfectly recovered, with a score of 96%. The fact that EE choices are less recovered is
 1024 not surprising, as the E-option estimates results from the comparison of one option
 1025 against 7 others, when in the ES phase an E-option is presented against a wider range
 1026 of alternative options (11), hence allowing better precision in the fitting of E-option value
 1027 estimates.

1028 We then generate new E-option value estimates, by applying our initial logistic fitting
 1029 procedure (see methods section) to this newly simulated data. We observe that E-option
 1030 value estimates are almost perfectly recovered, both in the ES (**Fig. S2**) and EE (**Fig. S3**)
 1031 phase, with a spearman ρ that is systematically higher than .97.

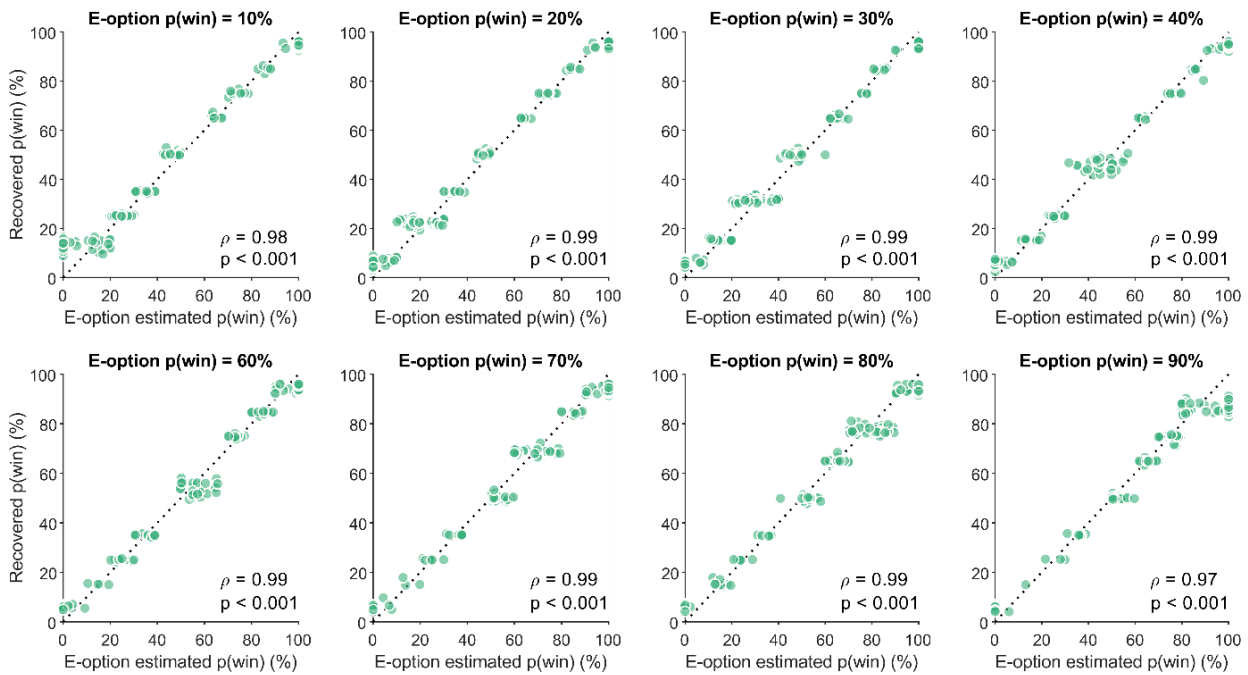


1032 .

1033 **Fig. S2 Recovery of E-option estimated probabilities in Experiments 1-to-6, ES phase.** We estimate 8 E-option
 1034 value for each subject in the ES phase. Thereafter, going through each individual choice history, we simulate a new
 1035 choice dataset using these value estimates as an input for an argmax decision rule. We apply our logistic fitting
 1036 procedure again (see the methods section) on this simulated data, to generate new estimates. Then we run a spearman
 1037 correlation to test the relationship between the estimates from the behavioral data and the estimates from the simulated
 1038 data. The grey dotted line corresponds to a perfect recovery of E-option probability estimates.

1039

1040



1041

1042 **Fig S3. Recovery of E-option estimated probabilities in Experiments 1-to-6, EE phase.** We estimate 8 E-option
 1043 value for each subject in the EE phase. Thereafter, going through each individual choice history, we simulate a new
 1044 choice dataset using these value estimates as input for an argmax decision rule. We apply our logistic fitting procedure
 1045 again (see the methods section) on this simulated data, to generate new estimates. Then we run a spearman
 1046 correlation to test the relationship between the estimates from the behavioral data and the estimates from the simulated
 1047 data. The grey dotted line corresponds to a perfect recovery of E-option probability estimates.

1048

1049 **Supplementary results**

1050 **Experiment 8**

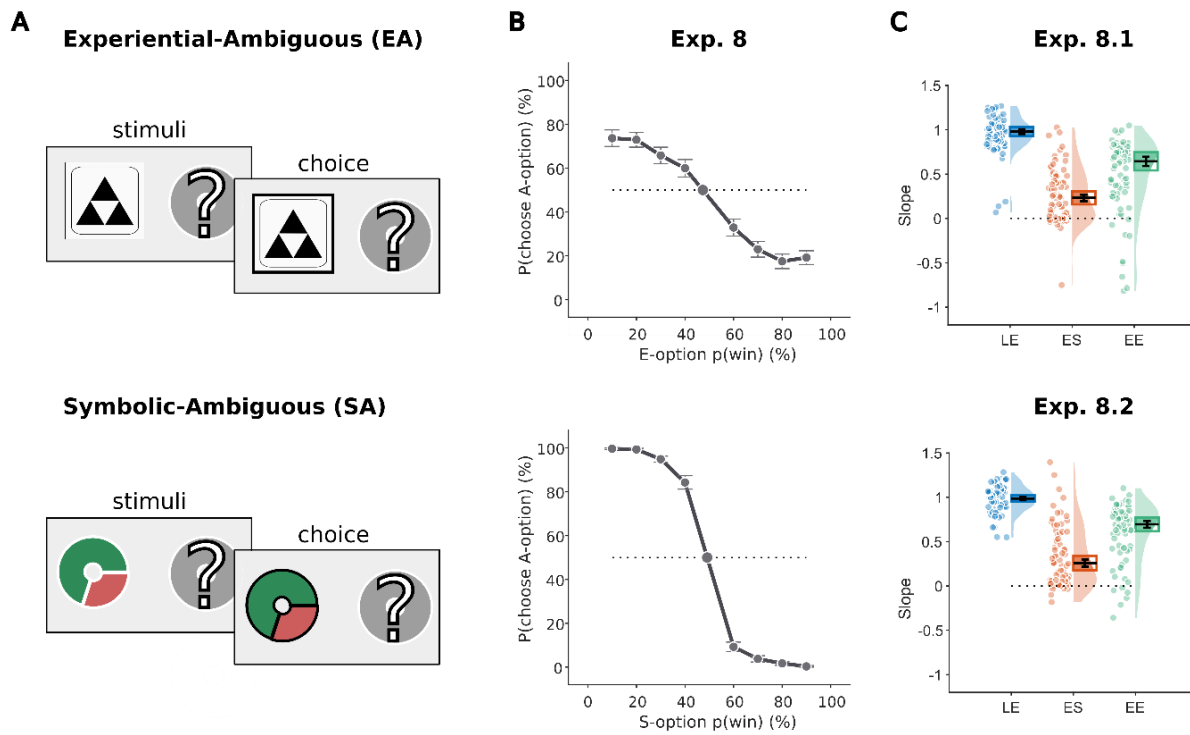
1051 We devised Experiment 8 to test whether the behavioral pattern observed in the
1052 Experiential-Symbolic phase was the result of ambiguity aversion (47), i.e. that the
1053 participant have a preference for options with known probability distributions over option
1054 with unknown probability distributions. In other words, participants would neglect
1055 experiential expected-values estimated during the LE phase because they are reluctant
1056 toward ambiguous options (E-options) and consequently mainly rely on options actually
1057 providing full probabilistic information (S-options).

1058 Thus, we presented each E-option (**Fig. S1A: top**) and S-option (**Fig. S1A: bottom**)
1059 against one ambiguous option (A-option), represented by a greyed pie-chart which
1060 conveyed no a priori information. Interestingly, the indifference point inferred for the A-
1061 option was close to 50% (both when the A-option is presented against E- and A-options).
1062 It suggests that without a priori information, participants associate a 50% subjective
1063 probability of winning a point to the A-option. When presented against E-options (**Fig.**
1064 **S1B: top**), the A-option is preferred against E-option which probability of winning a point
1065 is inferior to 50%, which suggests that those options are remembered as giving a negative
1066 expected-value. The preference is reversed when the E-option probability is above 50%,
1067 showing that participants associate those options to positive expected-values. When
1068 presented against S-options (**Fig. S1B: bottom**)

1069 Of note, E-options cannot be conflated with A-options for two reasons. First, when
1070 presented against A-options, E-options choice frequency increases monotonically with
1071 their associated objective probabilities. This result suggests that E-options are robustly
1072 linked to past outcome information, when it comes to comparing them to ambiguous
1073 stimuli.

1074 Second, the *experiential neglect* pattern cannot be the result of pure ambiguity aversion,
1075 as E-options are in average preferred against S-options when the latter has a negative
1076 expected-value, regardless of the E-option value. It suggests that this preference for

1077 known risks only holds in the gain domain, which excludes a pure preference toward
 1078 known risks, i.e., a pure ambiguity aversion.



1079

1080 **Fig. S1 Raw behavioral results and inferred option values in Experiments 8.** (A) The topmost panel
 1081 displays successive screens of a typical trial in the Experiential-Ambiguous (EA) phase. The bottommost
 1082 panel displays successive screens of a typical trials in the Symbolic-Ambiguous (SA) phase. The EA-phase
 1083 consists in binary choices between a symbol previous encountered in the LE-phase, and an ambiguous
 1084 lottery (materialized as greyed pie-chart with a question mark on top). The SA-phase consists in binary
 1085 choices between an explicit lottery (materialized as a pie-chart partly green for gain probabilities, and partly
 1086 red for loss probabilities) and an ambiguous lottery (materialized as greyed pie-chart with a question mark
 1087 on top). (B) Average probability of choosing an ambiguous option (A-option) over a E-option (top) or an S-
 1088 option (bottom) during the ambiguity phase. Dots represent the empirical choice frequency of the A-option.
 1089 The largest dot at the intersection of the grey dotted line represents the indifference point, i.e., when the
 1090 subject chooses randomly between the two options. The error bars represent the standard error of the
 1091 mean. (C) Comparison of individual inferred slopes obtained from linear fit in three modalities (LE, ES and
 1092 EE in blue, orange and green, respectively). The black lines represent mean and standard error of the
 1093 mean. The colored boxes represent 95% confidence interval. The shaded area represents the probability
 1094 density function. *** $p < 0.001$ paired sample t-tests.

1095

1096 Of note, introducing A-options among the other post-learning assessments did not affect
 1097 the previously observed relation in inferred slopes (Fig. S1C). LE-inferred slopes were
 1098 consistently significantly higher than ES slopes (Exp. 8.1: $T(72)=13.05$, $P < 0.001$; Exp

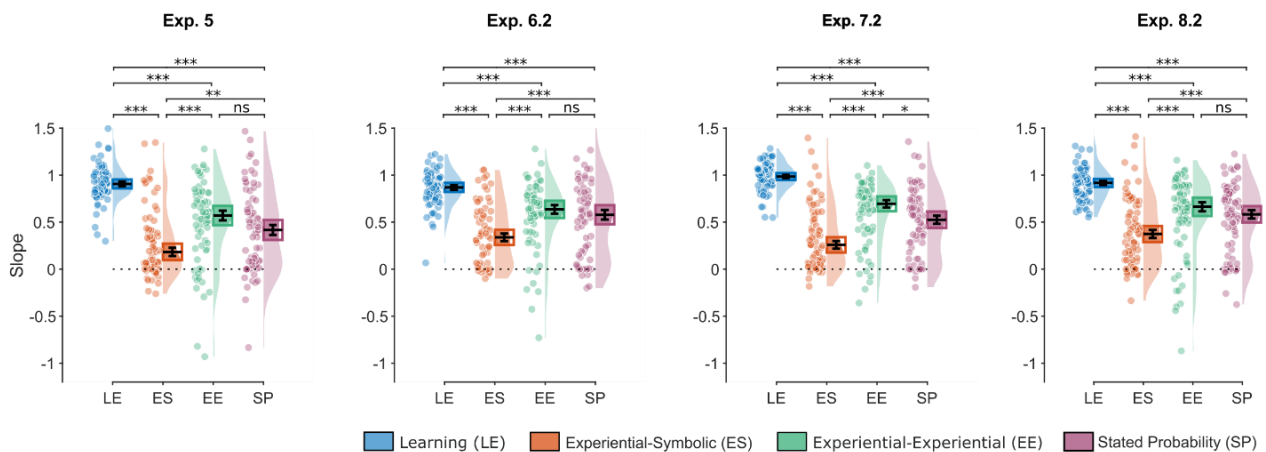
1099 **8.2:** $T(72)=10.41, P < 0.001$) as well as EE slopes (**Exp. 8.1:** $T(72)=5.9, P < 0.001$; **Exp**
 1100 **8.2:** $T(72)=6.38, P < 0.001$). EE-inferred slopes for their part were systematically higher
 1101 than ES slopes (**Exp. 8.1:** $T(72)=5.78, P < 0.001$; **Exp 8.2:** $T(72)=4.61, P < 0.001$).

1102 **Slope comparison among Experiments 5-to-8.**

1103 From experiments 5-to-8 (**Fig. S4**), LE slopes were consistently and significantly higher
 1104 than ES slopes (**Exp. 5:** $T(70)=12.94, P<0.001$; **Exp. 6.2:** $T(65)=10.59, P<0.001$; **Exp.**
 1105 **7.2:** $T(70)=14.4, P<0.001$;**Exp. 8.2:** $T(72)=10.41, P<0.001$), EE slopes (**Exp. 5:**
 1106 $T(70)=7.7, P<0.001$; **Exp. 6.2:** $T(65)=5.72, P<0.001$; **Exp. 7.2:** $T(70)=8.18, P<0.001$;**Exp.**
 1107 **8.2:** $T(72)=6.38, P<0.001$), as well as SP (**Exp. 5:** $T(70)=8.98, P<0.001$; **Exp. 6.2:**
 1108 $T(65)=6.18, P<0.001$; **Exp. 7.2:** $T(70)=10.88, P<0.001$;**Exp. 8.2:** $T(72)=7.71, P<0.001$).

1109 The EE slopes were the second closest to 1, i.e., the second closest to E-option objective
 1110 values. They are consistently higher than ES slopes (**Exp. 5:** $T(70)=4.48, P<0.001$; **Exp.**
 1111 **6.2:** $T(65)=4.84, P<0.001$; **Exp. 7.2:** $T(70)=7.77, P<0.001$;**Exp. 8.2:** $T(72)=4.61,$
 1112 $P<0.001$), however they are most of the time not significantly different from SP slopes
 1113 (**Exp. 5:** $T(70)=1.75, P=1$; **Exp. 6.2:** $T(65)=1.12, P=1$; **Exp. 7.2:** $T(70)=3.3, P<0.05$;**Exp.**
 1114 **8.2:** $T(72)=1.12, P=1$). SP slopes for their part, are systematically higher than ES slopes
 1115 (**Exp. 5:** $T(70)=4.05, P<0.01$; **Exp. 6.2:** $T(65)=5.34, P<0.001$; **Exp. 7.2:** $T(70)=4.8,$
 1116 $P<0.001$;**Exp. 8.2:** $T(72)=4.62, P<0.001$), designating the ES values as the lowest
 1117 slopes and the closest to 0.

1118



1119

1120 **Fig S4. Inferred option values in Experiments 5-to-8.** Comparison of individual inferred slopes obtained
1121 from linear fit in the 4 modalities (LE, ES, EE and SP in blue, orange, green and purple, respectively). The
1122 black lines represent mean and standard error of the mean. The colored boxes represent 95% confidence
1123 interval. The shaded area represents the probability density function. *** $p < 0.001$ paired sample t-tests.

1124

1125 **Choice profiling among Experiment 1-to-8**

1126 We classified ES-choices in different categories as a function of being explained
1127 exclusively either by a full E-value neglect, by E-option estimates elicited in the LE phase,
1128 by both, or finally by none of them (**Fig. S5**).

1129 In order to do so, we ran two simulations for each experiment.

1130 In the first simulation, for each subject, we simulate an agent that is confronted with the
1131 history of decision problems that the real subject was facing. This artificial agent makes
1132 decisions according to the following *experiential neglect* decision rule: If the S-option is
1133 above 50% chance of winning a point, choose the S-option, otherwise choose the E-
1134 option. This behavior is what we name *experiential neglect*, because the values of the E-
1135 options are not even considered by the decision-maker.

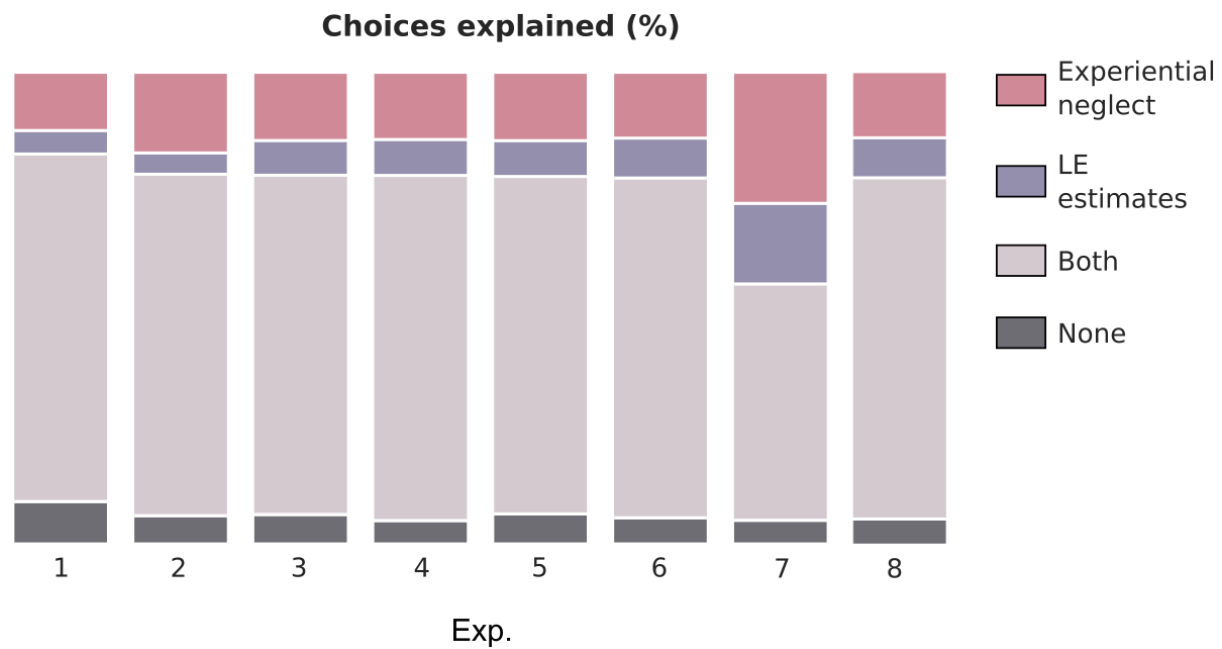
1136 In the second simulation, we also simulate an agent that is confronted with the history of
1137 decision problems that the real subject was facing. However, this agent has access to the
1138 E-option value estimates (specific to the subject in question) that were inferred from the
1139 LE-phase through our Q-learning model fitting procedure (see methods). We do not
1140 assume any deformation regarding the perception of S-option probabilities and rewards.
1141 Consequently, we assume that the agent uses an argmax rule (i.e., systematically
1142 choosing the highest value), to decide between the (subjective) E-option estimates and
1143 the S-option objective expected-value.

1144 With this simulated choice dataset, we can compute the proportion of choices from our
1145 behavioral data that match with each simulation.

1146 We observe that most of the choices can be both explained by LE estimates and the
1147 extreme *experiential neglect* decision rule (see main text). Yet, the number of choices

1148 exclusively explained (or predicted) by the *experiential neglect* rule is significantly higher
 1149 than the number of choices explained by LE estimates ($T(598)=13.87, P<0.001$).

1150 Of note, the Exp. 7, due to its particular configuration of probabilities among E- and S-
 1151 options, seemingly allows to discriminate better between the two decision models, as the
 1152 number of choices explained by both decision rules.



1153 **Fig S5. Choices prediction from Experiments 1-to-8.** We run 2 simulations. In the first one, we assume
 1154 that all participants make use of an *experiential neglect* decision rule, which basically consists in choosing
 1155 the S-option as long as it is higher than 50% chance of winning a point, and otherwise choose the E-option.
 1156 The second one consists in simulating all choices while assuming that participants use experiential values
 1157 from the LE phase (i.e., the ones we inferred through our Q-learning model), therefore “LE estimates”.
 1158 Thereafter we compute the proportion of choices that are explained by each simulation, i.e., the proportion
 1159 of behavioral choice that are identical to simulated choices. Choices explained by experiential neglect are
 1160 in red. Choices explained by inferred experiential values from the LE phase are in dark blue. Choices
 1161 explained by both experiential values and experiential neglect are in grey. Choices explained by none of
 1162 them are in black.
 1163

1164

1165

1166 **References**

- 1167 1. P. A. Samuelson, A Note on the Pure Theory of Consumer's Behaviour. *Economica*. **5**, 61–
1168 71 (1938).
- 1169 2. J. Von Neumann, O. Morgenstern, *Theory of games and economic behavior* (Princeton
1170 University Press, Princeton, NJ, US, 1944), *Theory of games and economic behavior*.
- 1171 3. A. Rangel, C. Camerer, P. R. Montague, A framework for studying the neurobiology of
1172 value-based decision making. *Nat. Rev. Neurosci.* **9**, 545–556 (2008).
- 1173 4. R. J. Herrnstein, Relative and absolute strength of response as a function of frequency of
1174 reinforcement. *J. Exp. Anal. Behav.* **4**, 267 (1961).
- 1175 5. B. F. Skinner, *Science and human behavior* (Simon and Schuster, 1965).
- 1176 6. R. S. Sutton, A. G. Barto, *Reinforcement learning: An introduction* (MIT press, 2018).
- 1177 7. D. Bernoulli, Exposition of a New Theory on the Measurement of Risk. *Econometrica*. **22**,
1178 23–36 (1738).
- 1179 8. P.-S. Laplace, *Essai philosophique sur les probabilités* (H. Remy, 1829).
- 1180 9. P. Wakker, A. Tversky, An axiomatization of cumulative prospect theory. *J. Risk Uncertain.*
1181 **7**, 147–175 (1993).
- 1182 10. D. Kahneman, A. Tversky, in *Handbook of the fundamentals of financial decision making:*
1183 *Part I* (World Scientific, 2013), pp. 269–278.
- 1184 11. B. De Martino, D. Kumaran, B. Seymour, R. J. Dolan, Frames, biases, and rational
1185 decision-making in the human brain. *Science*. **313**, 684–687 (2006).
- 1186 12. P. W. Glimcher, *Foundations of neuroeconomic analysis* (OUP USA, 2011).
- 1187 13. C. F. Camerer, A review essay about Foundations of Neuroeconomic Analysis by Paul
1188 Glimcher. *J. Econ. Lit.* **51**, 1155–82 (2013).
- 1189 14. I. Vlaev, N. Chater, N. Stewart, G. D. A. Brown, Does the brain calculate value? *Trends*
1190 *Cogn. Sci.* **15**, 546–554 (2011).
- 1191 15. N. Stewart, EPS Prize Lecture: Decision by sampling: The role of the decision environment
1192 in risky choice. *Q. J. Exp. Psychol.* **62**, 1041–1062 (2009).
- 1193 16. I. Erev, E. Ert, O. Plonsky, D. Cohen, O. Cohen, From anomalies to forecasts: Toward a
1194 descriptive model of decisions under risk, under ambiguity, and from experience. *Psychol.*
1195 *Rev.* **124**, 369 (2017).
- 1196 17. O. Bartra, J. T. McGuire, J. W. Kable, The valuation system: a coordinate-based meta-
1197 analysis of BOLD fMRI experiments examining neural correlates of subjective value.
1198 *Neuroimage*. **76**, 412–427 (2013).

- 1199 18. J. Garrison, B. Erdeniz, J. Done, Prediction error in reinforcement learning: a meta-
1200 analysis of neuroimaging studies. *Neurosci. Biobehav. Rev.* **37**, 1297–1310 (2013).
- 1201 19. J. A. Clithero, A. Rangel, Informatic parcellation of the network involved in the computation
1202 of subjective value. *Soc. Cogn. Affect. Neurosci.* **9**, 1289–1302 (2014).
- 1203 20. E. Fouragnan, C. Retzler, M. G. Philiastides, Separate neural representations of prediction
1204 error valence and surprise: Evidence from an fMRI meta-analysis. *Hum. Brain Mapp.* **39**,
1205 2887–2906 (2018).
- 1206 21. R. Hertwig, I. Erev, The description–experience gap in risky choice. *Trends Cogn. Sci.* **13**,
1207 517–523 (2009).
- 1208 22. C. R. Madan, E. A. Ludvig, M. L. Spetch, The role of memory in distinguishing risky
1209 decisions from experience and description. *Q. J. Exp. Psychol.* **70**, 2048–2059 (2017).
- 1210 23. D. U. Wulff, M. Mergenthaler-Canseco, R. Hertwig, A meta-analytic review of two modes of
1211 learning and the description-experience gap. *Psychol. Bull.* **144**, 140 (2018).
- 1212 24. B. Garcia, F. Cerrotti, S. Palminteri, The description–experience gap: a challenge for the
1213 neuroeconomics of decision-making under uncertainty. *Philos. Trans. R. Soc. B.* **376**,
1214 20190665 (2021).
- 1215 25. D. Kellen, T. Pachur, R. Hertwig, How (in)variant are subjective representations of
1216 described and experienced risk and rewards? *Cognition.* **157**, 126–138 (2016).
- 1217 26. I. Erev, E. Ert, A. E. Roth, E. Haruvy, S. M. Herzog, R. Hau, R. Hertwig, T. Stewart, R.
1218 West, C. Lebiere, A choice prediction competition: Choices from experience and from
1219 description. *J. Behav. Decis. Mak.* **23**, 15–47 (2010).
- 1220 27. T. H. B. FitzGerald, B. Seymour, D. R. Bach, R. J. Dolan, Differentiable Neural Substrates
1221 for Learned and Described Value and Risk. *Curr. Biol.* **20**, 1823–1829 (2010).
- 1222 28. S. R. Heilbronner, B. Y. Hayden, The description-experience gap in risky choice in
1223 nonhuman primates. *Psychon. Bull. Rev.* **23**, 593–600 (2016).
- 1224 29. W. M. DuCharme, M. L. Donnell, Intrasubject comparison of four response modes for
1225 “subjective probability” assessment. *Organ. Behav. Hum. Perform.* **10**, 108–117 (1973).
- 1226 30. R. A. Rescorla, A theory of Pavlovian conditioning: Variations in the effectiveness of
1227 reinforcement and nonreinforcement. *Curr. Res. Theory*, 64–99 (1972).
- 1228 31. T. E. J. Behrens, M. W. Woolrich, M. E. Walton, M. F. S. Rushworth, Learning the value of
1229 information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
- 1230 32. A. G. E. Collins, The Tortoise and the Hare: Interactions between Reinforcement Learning
1231 and Working Memory. *J. Cogn. Neurosci.* **30**, 1422–1432 (2018).
- 1232 33. S. Palminteri, M. Khamassi, M. Joffily, G. Coricelli, Contextual modulation of value signals
1233 in reward and punishment learning. *Nat. Commun.* **6**, 1–14 (2015).

- 1234 34. S. Bavard, A. Rustichini, S. Palminteri, Two sides of the same coin: Beneficial and
1235 detrimental consequences of range adaptation in human reinforcement learning. *Sci. Adv.*
1236 **7**, eabe0340.
- 1237 35. G. M. Becker, M. H. DeGroot, J. Marschak, Measuring utility by a single-response
1238 sequential method. *Behav. Sci.* **9**, 226–232 (1964).
- 1239 36. J. Rieskamp, P. E. Otto, SSL: A Theory of How People Learn to Select Strategies. *J. Exp.*
1240 *Psychol. Gen.* **135**, 207–236 (2006).
- 1241 37. L. J. Savage, *The foundations of statistics* (Courier Corporation, 1972).
- 1242 38. B. L. Lipman, Information Processing and Bounded Rationality: A Survey. *Can. J. Econ.*
1243 *Rev. Can. Econ.* **28**, 42–67 (1995).
- 1244 39. V. M. Chase, R. Hertwig, G. Gigerenzer, Visions of rationality. *Trends Cogn. Sci.* **2**, 206–
1245 214 (1998).
- 1246 40. G. Gigerenzer, W. Gaissmaier, Heuristic decision making. *Annu. Rev. Psychol.* **62**, 451–
1247 482 (2011).
- 1248 41. B. Mackowiak, F. Matejka, M. Wiederholt, Rational inattention: A review (2021).
- 1249 42. H. A. Simon, Theories of bounded rationality. *Decis. Organ.* **1**, 161–176 (1972).
- 1250 43. H. A. Simon, A. Newell, Human problem solving: The state of the theory in 1970. *Am.*
1251 *Psychol.* **26**, 145 (1971).
- 1252 44. G. E. Gigerenzer, R. E. Hertwig, T. E. Pachur, *Heuristics: The foundations of adaptive*
1253 *behavior*. (Oxford University Press, 2011).
- 1254 45. H. A. Simon, *Administrative behavior* (Simon and Schuster, 2013).
- 1255 46. D. Ellsberg, Risk, Ambiguity, and the Savage Axioms. *Q. J. Econ.* **75**, 643–669 (1961).
- 1256 47. D. Frisch, J. Baron, Ambiguity and rationality. *J. Behav. Decis. Mak.* **1**, 149–157 (1988).
- 1257 48. C. Camerer, M. Weber, Recent developments in modeling preferences: Uncertainty and
1258 ambiguity. *J. Risk Uncertain.* **5**, 325–370 (1992).
- 1259 49. I. Krajbich, B. Bartling, T. Hare, E. Fehr, Rethinking fast and slow based on a critique of
1260 reaction-time reverse inference. *Nat. Commun.* **6**, 7455 (2015).
- 1261 50. C. A. Holt, S. K. Laury, Risk aversion and incentive effects. *Am. Econ. Rev.* **92**, 1644–1655
1262 (2002).
- 1263 51. A. Tversky, D. Kahneman, Prospect theory: An analysis of decision under risk.
1264 *Econometrica.* **47**, 263–291 (1979).
- 1265 52. A. Tversky, D. Kahneman, Advances in prospect theory: Cumulative representation of
1266 uncertainty. *J. Risk Uncertain.* **5**, 297–323 (1992).

- 1267 53. D. Prelec, The Probability Weighting Function. *Econometrica*. **66**, 497 (1998).
- 1268 54. J. Quiggin, *Generalized expected utility theory: The rank-dependent model* (Springer
1269 Science & Business Media, 2012).
- 1270 55. A. Soltani, E. Koechlin, Computational models of adaptive behavior and prefrontal cortex.
1271 *Neuropsychopharmacology*, 1–14 (2021).
- 1272 56. K. Rayner, Eye movements in reading and information processing: 20 years of research.
1273 *Psychol. Bull.* **124**, 372 (1998).
- 1274 57. A. Glöckner, A.-K. Herbold, An eye-tracking study on information processing in risky
1275 decisions: Evidence for compensatory strategies based on automatic processes. *J. Behav.*
1276 *Decis. Mak.* **24**, 71–98 (2011).
- 1277 58. S. Fiedler, A. Glöckner, The Dynamics of Decision Making in Risky Choice: An Eye-
1278 Tracking Analysis. *Front. Psychol.* **3**, 335 (2012).
- 1279 59. V. Venkatraman, J. W. Payne, S. A. Huettel, An overall probability of winning heuristic for
1280 complex risky decisions: Choice and eye fixation evidence. *Organ. Behav. Hum. Decis.*
1281 *Process.* **125**, 73–87 (2014).
- 1282 60. J. A. Aimone, S. Ball, B. King-Casas, It's not what you see but how you see it: Using eye-
1283 tracking to study the risky decision-making process. *J. Neurosci. Psychol. Econ.* **9**, 137–
1284 144 (2016).
- 1285 61. R. S. Sutton, Learning to predict by the methods of temporal differences. *Mach. Learn.* **3**,
1286 9–44 (1988).
- 1287 62. P. Dayan, L. F. Abbott, *Theoretical neuroscience: computational and mathematical*
1288 *modeling of neural systems* (Computational Neuroscience Series, 2001).
- 1289 63. J. Li, N. D. Daw, Signals in Human Striatum Are Appropriate for Policy Update Rather than
1290 Value Prediction. *J. Neurosci.* **31**, 5504–5511 (2011).
- 1291 64. B. Y. Hayden, Y. Niv, The case against economic values in the orbitofrontal cortex (or
1292 anywhere else in the brain). *Behav. Neurosci.* **135**, 192 (2021).
- 1293 65. D. Bennett, Y. Niv, A. J. Langdon, Value-free reinforcement learning: policy optimization as
1294 a minimal model of operant behavior. *Curr. Opin. Behav. Sci.* **41**, 114–121 (2021).
- 1295 66. M. J. Frank, C. D'Lauro, T. Curran, Cross-task individual differences in error processing:
1296 neural, electrophysiological, and genetic components. *Cogn. Affect. Behav. Neurosci.* **7**,
1297 297–308 (2007).
- 1298 67. A. G. Collins, M. J. Frank, Opponent actor learning (OpAL): modeling interactive effects of
1299 striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev.* **121**, 337
1300 (2014).
- 1301 68. M. Möller, R. Bogacz, Learning the payoffs and costs of actions. *PLOS Comput. Biol.* **15**,
1302 e1006285 (2019).

- 1303 69. P. Redgrave, T. J. Prescott, K. Gurney, The basal ganglia: a vertebrate solution to the
1304 selection problem? *Neuroscience*. **89**, 1009–1023 (1999).
- 1305 70. I. Bar-Gad, H. Bergman, Stepping out of the box: information processing in the neural
1306 networks of the basal ganglia. *Curr. Opin. Neurobiol.* **11**, 689–695 (2001).
- 1307 71. L. T. Hunt, N. Kolling, A. Soltani, M. W. Woolrich, M. F. Rushworth, T. E. Behrens,
1308 Mechanisms underlying cortical activity during value-guided choice. *Nat. Neurosci.* **15**,
1309 470–476 (2012).
- 1310 72. A. Rustichini, C. Padoa-Schioppa, A neuro-computational model of economic decisions. *J.*
1311 *Neurophysiol.* **114**, 1382–1398 (2015).
- 1312 73. C. Padoa-Schioppa, K. E. Conen, Orbitofrontal cortex: a neural circuit for economic
1313 decisions. *Neuron*. **96**, 736–754 (2017).
- 1314 74. S. Farashahi, C. H. Donahue, P. Khorsand, H. Seo, D. Lee, A. Soltani, Metaplasticity as a
1315 Neural Substrate for Adaptive Learning and Choice under Uncertainty. *Neuron*. **94**, 401-
1316 414.e6 (2017).
- 1317 75. A. C. Huk, M. N. Shadlen, Neural Activity in Macaque Parietal Cortex Reflects Temporal
1318 Integration of Visual Motion Signals during Perceptual Decision Making. *J. Neurosci.* **25**,
1319 10420–10436 (2005).
- 1320 76. C. A. Hutcherson, B. Bushong, A. Rangel, A neurocomputational model of altruistic choice
1321 and its implications. *Neuron*. **87**, 451–462 (2015).
- 1322 77. C. Findling, N. Chopin, E. Koechlin, Imprecise neural computations as a source of adaptive
1323 behaviour in volatile environments. *Nat. Hum. Behav.* **5**, 99–112 (2021).
- 1324 78. R. Bricet, “Preferences for information precision under ambiguity” (THEMA (THéorie
1325 Economique, Modélisation et Applications), Université de ..., 2018).
- 1326 79. R. Dukas, Costs of Memory: Ideas and Predictions. *J. Theor. Biol.* **197**, 41–50 (1999).
- 1327 80. H. Afrouzi, S. Kwon, Y. Ma, “A Model of Costly Recall” (working paper, Columbia
1328 University, 2020).
- 1329 81. K. J. Miller, A. Shenhav, E. A. Ludvig, Habits without values. *Psychol. Rev.* **126**, 292
1330 (2019).
- 1331 82. I. Krajbich, C. Armel, A. Rangel, Visual fixations and the computation and comparison of
1332 value in simple choice. *Nat. Neurosci.* **13**, 1292–1298 (2010).
- 1333 83. P. Sepulveda, M. Usher, N. Davies, A. A. Benson, P. Ortoleva, B. De Martino, Visual
1334 attention modulates the integration of goal-relevant evidence and not value. *eLife*. **9**,
1335 e60705 (2020).
- 1336 84. S. Lichtenstein, P. Slovic, *The construction of preference* (Cambridge University Press,
1337 2006).

- 1338 85. B. Hayden, Y. Niv, The case against economic values in the brain (2020),
1339 doi:10.31234/osf.io/7hgup.
- 1340 86. R. D. Luce, The choice axiom after twenty years. *J. Math. Psychol.* **15**, 215–233 (1977).
- 1341 87. R. D. Luce, *Individual choice behavior: A theoretical analysis* (Courier Corporation, 2012).
- 1342 88. M. Lebreton, K. Bacily, S. Palminteri, J. B. Engelmann, Contextual influence on confidence
1343 judgments in human reinforcement learning. *PLoS Comput. Biol.* **15**, e1006973 (2019).
- 1344 89. R. C. Wilson, A. G. Collins, Ten simple rules for the computational modeling of behavioral
1345 data. *eLife.* **8**, e49547 (2019).
- 1346
- 1347

Entre le phénomène scientifique et le noumène scientifique, il ne s'agit donc plus d'une dialectique lointaine et oisive, mais d'un mouvement alternatif qui, après quelques rectifications des projets, tend toujours à une réalisation effective du noumène. Elle renforce ce qui transparait derrière ce qui apparaît. Elle s'instruit par ce qu'elle construit.

Gaston Bachelard, *Le Nouvel Esprit Scientifique*, 1934

6

Discussion

In chapter 1, we saw that utility models of economic decisions derive from the will to quantify economic value. The ontology of value (i.e. its modalities of existence) as well as its epistemology (i.e. how to study it) have varied over time. At first its conception was materialist and objective. The labor theory value thus presupposes that economic value is embodied in a good. Subsequently, subjective conceptions of value (utility) were imposed in the academic field, notably via neoclassical theories. The axiomatic utility model was gradually stripped of all psychological considerations. When these axioms were refuted on empirical grounds, a battle ensued over the normative model to be used to describe economic decisions. Some defend an approach based on subjective and empirically characterized (and possibly neurally implemented) models of value (prospect theory for

instance). These models are mostly rooted in the description paradigm, i.e. the probabilities and rewards of options are described *a priori*. This way of operationalizing the decision stems from the assumptions of rationality, notably the assumptions that economics agent have access to full information. Other researchers consider that the heuristics and value-free approaches are more appropriate, as rationality is bounded and ecological. They argue that heuristics allow to make efficient decisions in complex and information lacking environments.

In chapter 2, we saw that RL models emerged within the paradigm of classical conditioning. Behaviorists considered that psychological phenomena should be reduced to a behavioral output-input model, and were mostly focused on animal learning. However, a more cognitivist approach developed when computational models of classical conditioning were applied to human decision-making. The experience paradigm of decision-making was greatly inspired by animal learning. Notably, humans are confronted to multi-armed bandit task where information is lacking, and the contingencies of their actions are learned through trial-and-error. The value-functions, the decision rules, as well as their parameters, took an important role in modeling and describing human learning and decisions. Also, the field of value-based decision making definitely established itself when evidences for material translations of RL components (e.g. prediction-error) were found in the brain. On this basis, the *neural common currency hypothesis* emerged. It specifies that there exists a dedicated cortical circuit for valuation, which allows the comparison of items that fundamentally differ in nature through a common neural representation. Afterwards, the two-step model (unified valuation of options, then selection of the option maximizing expected value) of value-based decision making became more and more dominant. In the mean time, and for the sake of parsimony (i.e. not invoking the value construct), other scholars have fostered alternative pathways to describe human decisions, notably via policy models (e.g. policy gradient learning).

In chapter 3, we saw that the meeting of the description and experience paradigms resulted in behavioral discrepancies, also known as the *description-experience gap*. More precisely, the sub-

jective valuation of described and experienced probabilities and outcomes seemingly differs. We discussed the literature on the possible determinants of this behavioral gap.

In chapter 4, we studied the implications of the *description-experience gap* for cross-species studies of decision under uncertainty. We concluded that the rhesus monkey only constitutes a partial model of human decision-making under uncertainty. Indeed, in the description domain, when humans are typically risk-averse in gains and risk-seeking in losses, monkeys often display opposite preferences. However, in the experience domain, humans and monkeys display similar risk-attitudes. One explanation could be that, when description-based studies in monkeys require learning a symbolic system from scratch, in humans the meaning of risk is provided by language. As a result, monkeys are located in a mixed paradigm (i.e. 'description + experience'), where they have to learn symbolic options through trial-and-error, i.e. by experience. Additionally, we identified several methodological gaps (such as the nature of the reward, or the number of trials) which might prevent from proper cross-species comparisons. We proposed further lines of enquiry that could help reducing these gaps, and foster methodological overlaps between humans and non-human primates decision-making study.

In chapter 5, we tested in a human behavioral study, the degree of commensurability of experiential versus symbolic options, and therefore how the *description-experience gap* is instantiated in this kind of hybrid decisions. Actively learning subjective values through experience does not entail the ability to properly compare those values to described symbolic ones. Despite subjects displaying high performance during the learning phase, the comparison of experiential and symbolic options are made almost regardless of the experiential values. We named this phenomenon *experiential neglect*. With various controls, we made sure that subjects do not merely forget experiential values. Thereafter, when presenting experiential options against each other, we observed that estimates were more in line with objective values, showing that this phenomenon does not result from an incapacity to extrapolate values to new decisions contexts. Moreover, even when

significantly modifying the configuration of probabilities attributed to experiential and symbolic options, subjects persist in the exact same behavior, despite that it comes at a significant economic cost. For this reason, we suggested that experiential and symbolic values are not only constructed and conveyed in two different ways, but also possibly rely on different representational systems.

However, this study presents several limitations. For instance, we aimed at testing (only behaviorally) the impact of the *description-experience gap*, and especially hybrid choices, on the traditional two-step model of decision making. This model specifies a unified valuation stage followed by a choice stage. Consequently, we hypothesized that experiential and symbolic values are built and retrieved in order to be further compared. Nevertheless, subjects' behavior in this experiment could be interpreted within a value-free framework as well, i.e. without assuming value computation, and therefore, value representations. Additionally, we have superficially considered the role of *ambiguity aversion* in our results. In fact, the behavioral pattern observed in our subjects could derive from both *ambiguity aversion* and *experiential neglect*.

In this discussion, we will address the various implications of the study presented in chapter 5, with respect to the contemporary literature.

6.1 Experiential and symbolic hybrid choices in previous literature

6.1.1 In monkeys

Heilbronner and Hayden (2016) presented hybrid choices between experienced and described options in monkeys. Three adult male rhesus macaques (labeled B, J, K), had to perform these hybrid choices by comparing explicit lotteries represented by colored bars, to abstract experienced cues (Fig. 6.1A).

In their task, they observed an overall preference for experienced options. Considering the entire

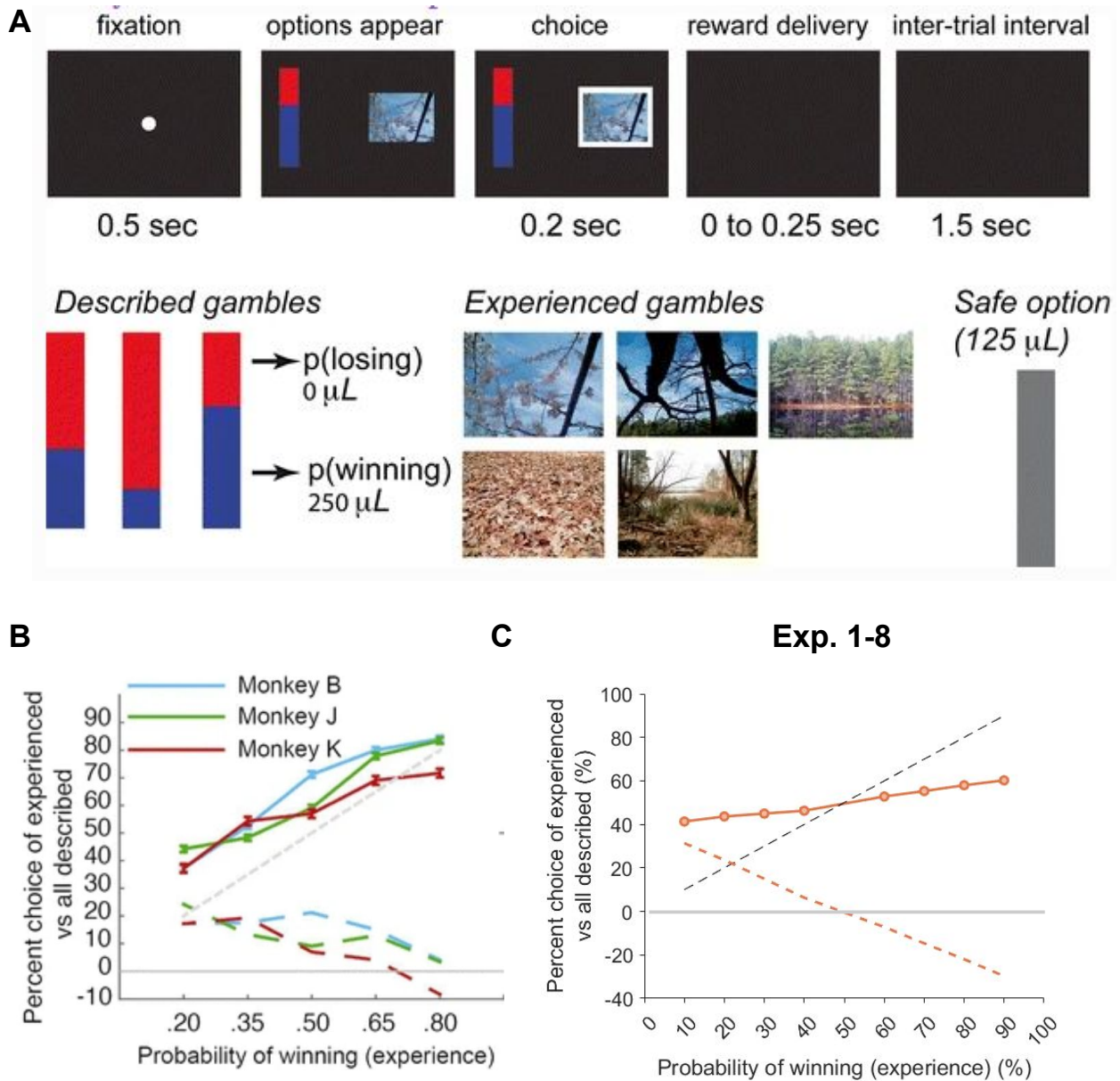


Figure 6.1: Experiential and symbolic hybrid choices in monkeys. **(A)** Three adult male rhesus macaques (subjects B, J, K), on each trial, had to choose between two options, a described and an experienced gamble, by shifting gaze to it and maintaining that gaze for 200 ms. Experiential options were five emotionally neutral nature scenes. Each scene corresponded to a win probability of respectively 20%, 35%, 50%, 65%, and 80%. Conversely described symbolic options were divided into a red and a blue portion, and indicated the probability of respectively losing and winning. Monkeys received water as a reward or punishment, respectively $250 \mu\text{L}$ or $0 \mu\text{L}$. **(B)** Probability of choosing the experiential option, according to probability of winning, presented against all symbolic options, for each monkey. Dashed colored lines indicate results from each subject minus description-experience neutrality (gray dashed line). **(C)** Probability of choosing the experiential option, according to probability of winning, presented against all symbolic options. All experiments (except from 4 and 7, which experiential options had different win probabilities) were pooled. Dashed colored lines indicate results from each subject minus description-experience neutrality (gray dashed line). A and B are from Heilbronner and Hayden. 2016

set of experiential options, monkeys chose the experienced over the range of described options more than half the time as well (B: 65.01%, J: 62.58%, K: 58.06%). It has been previously shown that humans prefer risky options to ambiguous ones (Curley et al., 1986; Einhorn and Hogarth, 1985) as well as monkeys (Hayden et al., 2010). One could suggest that because experiential options are more ambiguous they should be avoided. However, monkeys here displayed an opposite preference. When taking each experiential cue individually and ranking them according to probability of winning, we can observe a monotonic increase in their choice frequency coupled to a global overestimation (Fig. 6.1B).

In contrast, our human subjects display no particular preference toward one modality over the other. In fact, averaged over experiments 1-to-8 (with exp. 4 and 7 excluded, because they have different options' probabilities), subjects chose the experienced option precisely 50% of the time. Also, experiential options were over selected in losses (i.e., when the option is below 50% chance of winning a point) and under selected in gains (i.e., when the option is above 50% chance of winning a point) (Fig. 6.1C).

What could explain this gap between species? Unlike humans, and as shown in chapter 4, monkeys are reliably risk-seeking (Heilbronner and Hayden, 2013; Xu and Kralik, 2014). This study suggests that monkeys are even more risk-seeking for experienced cues. However, it is hard to disentangle experience from description learning in monkeys, as all subjects had extensive prior training (thousands of trials) to learn the symbolic system of described gambles, while the experiential gambles were newly learned for the study. Thus, this prior training could explain why there exists such a gap between species: value representations of experiential and symbolic options are constructed in a similar way in monkeys. In addition, losses in monkeys are hard to implement experimentally, and consequently a punishment consists in an absence of gains. This constraint might result in a shift of the reference point, and therefore the absence of asymmetry between the gain of loss domain, contrary to our task where our subjects displayed a kind of *reflection effect*.

Consequently, this difference between species in the comparison of experiential against symbolic options, calls for additional comparative studies. A way to make between species comparison possible, would be to study the experienced gain domain more precisely in humans. If the reference point moves, and the *reflection effect* (i.e, the *experiential neglect*) pattern is maintained, then we will be in the presence of a robust cross-species behavioral gap.

6.1.2 In humans

To our knowledge, [FitzGerald et al. \(2010\)](#) is the only human study using a behavioral paradigm identical to ours. Seventeen subjects underwent an fMRI task, in which they made choices between three experiential cues (with probability of winning 10%, 50% and 90%) and nine symbolic cues (with probability of winning 5%, 10%, 20%, 40%, 50%, 60%, 80%, 90%, and 95%). Symbolic cues probabilities were described with pie-charts, when experiential cues were basic geometrical forms. Subjects received a total of 160 trials of feedback per experiential cue.

Interestingly, when looking at indifference points between experiential and symbolic options, subjects significantly overweighted the low probability experienced option (Fig. 6.2A). On the other hand, the 50% and 90% chance option were well estimated, with indifference points in line with their objective values.

In our results, low probability experiential options were also overweighted, yet their high probability counterpart were underweighted. A possible explanation, is then that our *experiential neglect* emerges in order to cope with a saturation of working memory, or is due to sampling error. Indeed, in our task (Fig. 6.2B), subject only experienced 30 trials per cue. However, in experiment 4, we reduced the number of options to four, showed the feedback of both options, and doubled the number of trials. Even by these standards, the *experiential neglect* pattern remained. Moreover, subjects showed that they were able to rank the experiential options correctly when the latter were pitted against each others. Further behavioral studies should investigate the minimal conditions under which experiential options can be rationally assessed against symbolic ones.

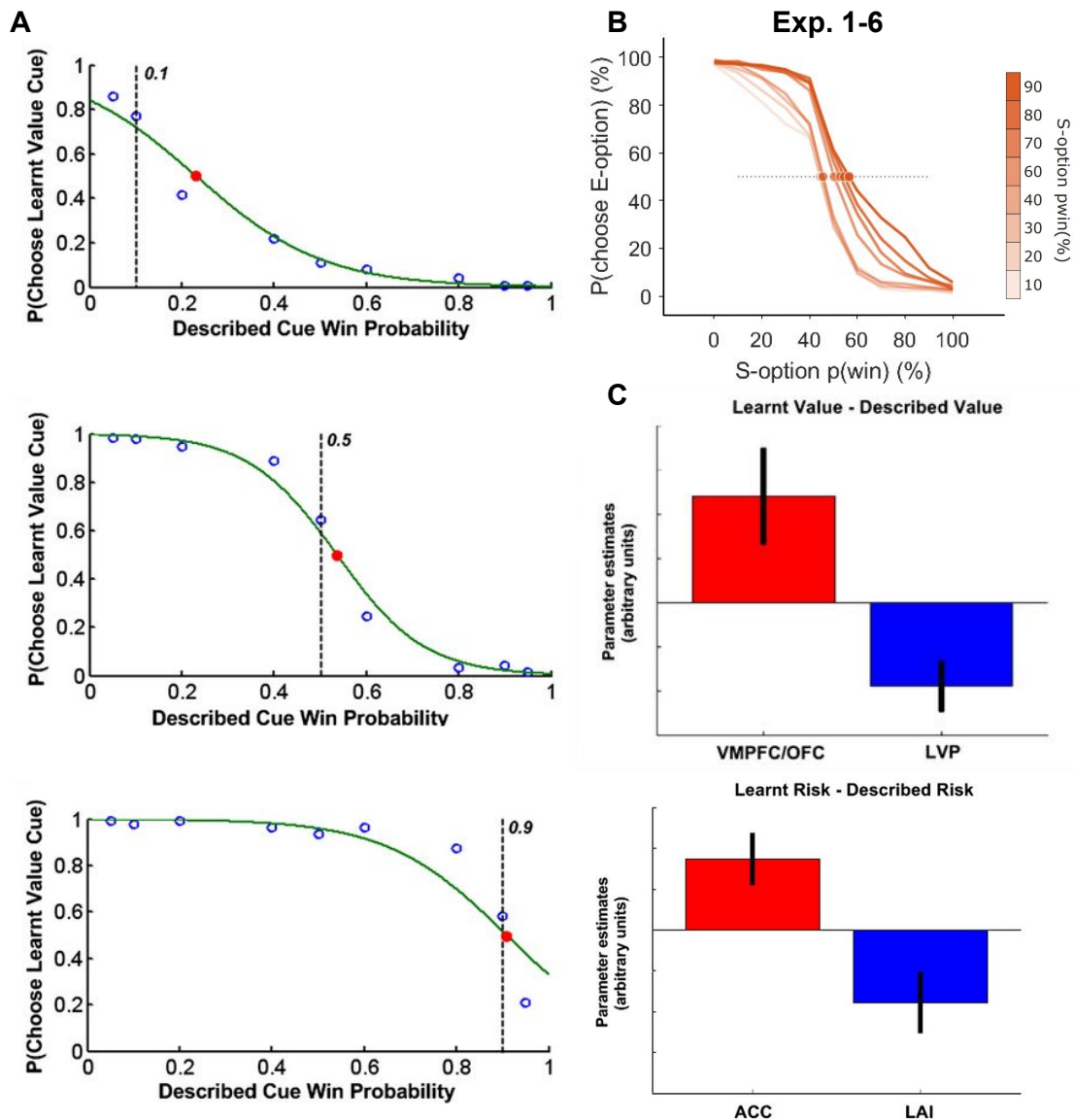


Figure 6.2: Experiential and symbolic hybrid choices in humans. **(A)** Indifference curves between an experiential option of probability 0.1 (top), 0.2 (middle), 0.3 (bottom), and a range of symbolic options ($p=0.05, 0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 0.9, 0.95$). The red dot represents the indifference points, while the blue dots presents the frequency for choosing the experiential option. The green curve is a logistic fit. **(B)** Indifference curves plotted for our experiments 1-to-6 (excluding experiment 4). Eight experiential options (E-options; $p=0.1, 0.2, 0.3, 0.4, 0.6, .7, 0.8, 0.9$) are presented against eleven symbolic options (S-options; $p=0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, .7, 0.8, 0.9, 1$). **(C)** (Top) Mean parameter estimates for activation in the ventromedial prefrontal cortex (vmPFC), medial orbito frontal cortex (mOFC) in red. Mean parameters estimates for activation in left ventral putamen (LVP). Activity in the vmPFC/OFC was correlated more strongly with the value of learned cues than described ones, whereas the left VP showed the opposite pattern. Black bars indicate 95% confidence. (Bottom) Mean parameter estimates for activation in the anterior cingulate cortex (ACC) in red. Mean parameters estimates for activation in left anterior insula (LAI). Activity in the ACC was correlated more strongly with the value of learned cues than described ones, whereas the LAI showed the opposite pattern. Black bars indicate 95% confidence.). A and C are from [FitzGerald et al., 2010](#).

Regarding the neural substrates involved (Fig. 6.2C), they found that vmPFC/OFC responded to experiential value, which fits with previous literature (e.g. Schoenbaum and Roesch, 2005; Padoa-Schioppa and Assad, 2006; Hare et al., 2008; Chib et al., 2009). Learned risk, i.e. the outcome variance, was correlated with the ACC. This phenomenon is also known as the *expected risk hypothesis* (Brown and Braver, 2005, 2008). Distinct neural substrates were found for symbolically described values. Notably, described risk was linked to the insula and described values were link to the ventral putamen. These structures are usually not attributed a prominent role in description-based choices (Padoa-Schioppa and Conen, 2017). Hence, these activations may be specific to hybrid choices.

At the neural level, further research could consists in identifying the brain structures involved in the processing of hybrid choices. For this we should uncover the regions involved in this kind of choices via behavioral fMRI monitored task in humans and build computational models that could account for these decisions. By relying on microscopic assessments and electrophysiological methods in monkeys, those models could be falsified and further refined.

6.2 General considerations on the idea of a representational gap

We hypothesized that our *experiential neglect* pattern was due to different value representations, and we discussed in chapter 5 the possible neural substrate underlying this process. The idea that subjective representations of experienced and described values systematically differ is more or more considered (e.g. Kellen et al., 2016). This *representational gap* that emerges from the comparison of experiential and symbolic values, might be an instantiation at the individual level of the *description-experience gap* observed at an aggregated level. Furthermore, this gap might be located at several description levels, and possibly with interactions between them:

- *Neural level*: Experiential and symbolic values might be implemented via different neural substrates, and consequently recruit different cognitive processes and mental representations. In chapter 5, we discussed how working memory could be specific to the building and retrieval of experiential values. Here we will explore this level by addressing the question of relative valuation. In addition, we will discuss how policy learning questions the relevance of postulating value representations, and therefore the traditional two-step model of value-based decision making.
- *Computational and action selection level*: Regardless of the valuation process and its implementation, the *representational gap* could be located at a higher cognitive level. For instance, our *experiential neglect* pattern could be caused by the hijacking of the action selection process by an alternative (and possibly more appropriate) decision rule. For that matter we will interpret our results through the lens of ambiguity aversion and heuristic decision-making.
- *Elicitation level*: It is possible that values are constructed at an even higher level of abstraction, namely in the process of elicitation. However, different elicitation methods often result in systematically different responses. We will discuss how the violation of procedure invariance might help understanding our results.

6.2.1 Are value representations relative?

In the last twenty-five years, a spectrum of neural and behavioral findings pointed out that valuation might be performed in a relative way (or at least that values are neurally rescaled according to the range of outcomes) in both description (Sugrue et al., 2004; Padoa-Schioppa, 2009) and experience domain, albeit particularly in experience (Tremblay and Schultz, 1999; Cromwell et al., 2005; Palminteri et al., 2015; Klein et al., 2017; Bavard et al., 2018, 2021; Isoda, 2021; Palminteri and Lebreton, 2021).

Indeed, in most daily life situations, decisions are contextual. For instance, choosing how to dress

for a particular event will depend on the social context in which that event takes place. This property of real-life decisions is well illustrated in perceptual decision-making (for reviews, see [Bar, 2004](#); [Schwartz et al., 2007](#)). For example, in the classic Ebbinghaus illusion, two circles of similar size are positioned near each other. Many larger circles are positioned around the central first one, while smaller circles are surrounding the other. As a consequence, people often perceive the circle surrounded by larger circles as smaller than the other one, suggesting that an object's subjective size perception is modulated by the properties of its surroundings, and more generally is influenced by relative judgments.

In the RL literature, growing evidence suggests that learning is sensitive to contextual effect (e.g., [Louie and De Martino, 2014](#); [Palminteri et al., 2015](#); [Bavard et al., 2018](#); [Palminteri and Lebreton, 2021](#)). This contextual learning has been identified to be generally located at two levels: reference-dependence and range-adaptation.

Reference-dependence refers to the valuation of gains and losses relative to a temporal or spatial reference point. The reference point is central to prospect theory ([Tversky and Kahneman, 1979](#)), as the *reflection effect* for instance, is articulated around it. For example, in loss-avoidance contexts, an avoided-loss may become rewarding (i.e. a relative reward) if losses were the most frequent outcome in the given context. In the experience domain, this reference-dependence has been shown to improve learning in losses, yet at the cost of irrational preferences, as learned values seemingly cannot be extrapolated to other decision problems ([Bavard et al., 2018, 2021](#)). Additionally, reference-dependence has also been characterized neurally in the description paradigm ([Weber et al., 2007](#)).

Studies investigating the contextual effects occurring in RL also highlighted the role of range-adaptation. At the behavioral level, subjects show different sensitivity for different ranges of intensity/magnitude ([Bavard et al., 2018](#)). Concomitantly and as a biological translation of these

findings, some evidence for neuronal range adaptation has been brought to light (Padoa-Schioppa, 2009). Neuronal range adaptation posits that the firing rate of neurons adapts to how the variable encoded is distributed, as a mechanism to cope with various decision situations.

Recent studies have shown that computational models combining a reference point and range-adaptation mechanism can explain various irrational preference patterns (Bavard et al., 2021). Interestingly enough, while ecological RL seems to be better accounted by models integrating a combination of reference-dependence and range-adaptation, the latter models perform poorly in description-based choices (Dumbalska et al., 2020; Landry and Webb, 2021).

Assuming experiential values formed in our learning phase are likely to be represented in a relative fashion, this phenomenon could account for our *experiential neglect* pattern to a certain extent. Indeed, if in the learning phase, relative values are learned, then it would prevent the comparison with absolute symbolic ones, and (potentially) induce the use of alternative decision rules in order to overcome the cost of comparing two incommensurable values. Hence, future research could test this hypothesis by designing a learning phase where experiential options are not learned by pairs, but rather presented against all other options, in order to reach a higher level of generalization.

6.2.2 Ambiguity aversion

While the term “ambiguity effect” was not coined in it, the underlying principles of it were already described in Ellsberg (1961) seminal paper. In this paper, Ellsberg outlines an hypothetical gamble:

Ellsberg’ paradox

- (1) You can win \$100 by drawing a ball of a certain color from a bucket.
- (2) The bucket contains 90 balls, 30 of which are red.
- (3) Among the 60 remaining balls, an unknown proportion are yellow, and the rest are black.
- (4) You have to bet \$100 either on a red ball, either a yellow ball.
- (5) Drawing a ball of the color you bet on will win you the \$100.
- (5) If you draw a black ball or a ball of the color you did not bet on, you will get nothing.

Ellsberg predicted that the majority would prefer to bet on the red ball ¹. The odds of drawing a red ball are $\frac{1}{3}$. Moreover, without further information, the prior for the proportion of yellow ball is also $\frac{1}{3}$. The proportion of red balls is consequently equal to the assumed proportion of yellow balls. The reason for this preference has therefore been interpreted in terms of 'ambiguity aversion', i.e., the preference for known risk rather than unknown risk. Said differently, most people avoid the option with missing outcome or probability information. In a variety of experimental and real-world situations, it has been shown that humans reliably prefer risky options to ambiguous ones, even paying an extra cost to avoid ambiguity (Einhorn and Hogarth, 1985; Curley et al., 1986; Camerer and Weber, 1992; Fox and Tversky, 1995). Likewise, monkeys also exhibit this preference (Hayden et al., 2010). Several explanations have been proposed for ambiguity aversion, such as comparative ignorance (Frisch and Baron, 1988; Fox and Tversky, 1995; Fox and Weber, 2002), i.e. the tendency to bet on what feels more familiar, or the belief that the ambiguous urn is rigged (Frisch and Baron, 1988; Kühberger and Perner, 2003). Yet, why people avoid ambiguity remains unclear.

Coming back to our *experiential neglect* result, the tendency for our subjects to choose symbolic and described options over experienced options could be interpreted as the result of ambiguity aversion, as experienced options are lacking symbolic information. For this reason, we included a control in our 8th experiment, where we presented one 100% ambiguous option against a set of previously encountered experiential and symbolic options. Interestingly, subjective estimates for the ambiguous option inferred from these choices were around 50%, i.e. subjects seemingly assign a neutral expected-value of 0 to the ambiguous option, thus closely resembling the *experiential neglect* pattern (see chapter 5, supplementary materials). Other studies have found a similar pattern, when presenting symbolic and risky options to ambiguous options (Li et al., 2015, 2017), which is pictured in Figure 6.3B/D.

¹His intuition will be verified by Halevy, 2007

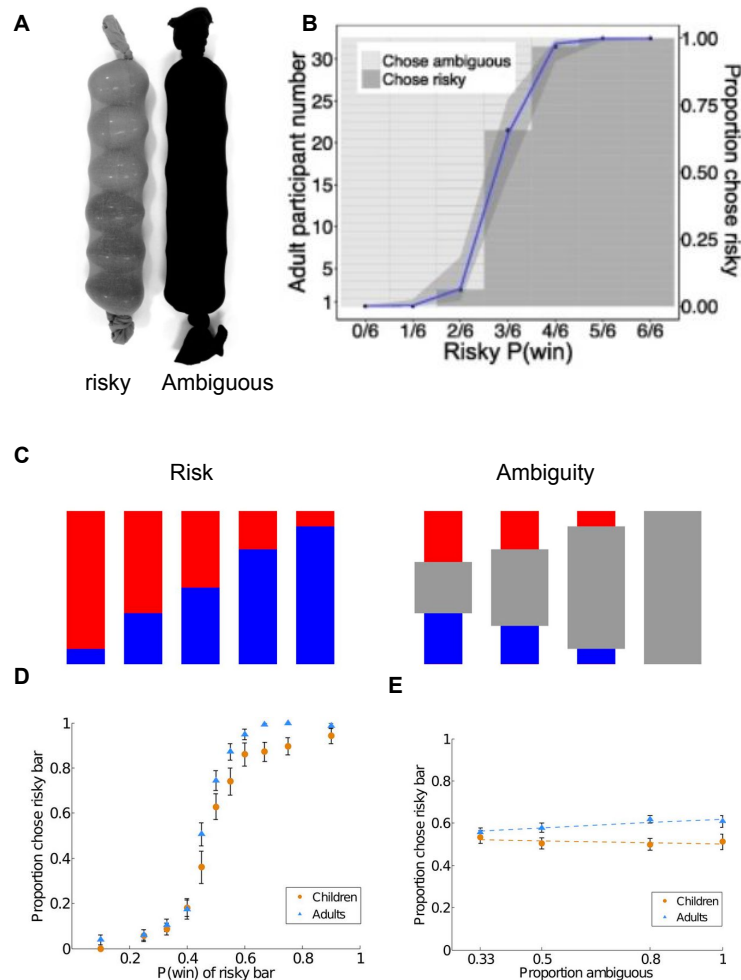


Figure 6.3: Ambiguity aversion pattern. (A) Risky versus ambiguous option. Subjects had to choose between physical stimuli representing risky gambles (in this example, 3 winning and 3 losing eggs; left) and ambiguous gambles (always containing 6 eggs of unknown colors; right). (B) The x-axis represents the probability of the risky option. The y-axis represents the probability of choosing the risky option. Each row of the shaded represents a participant, and each column a choice. When ambiguity is chosen, the cell is in light gray. When the risky option is chosen, the cell is in dark gray. The blue line is a logistic regression, that represents the proportion of risky choices according to its probability. (C) Risk and ambiguous options visual representations. Red proportions indicates the probability of losses, while blue indicates the probability for gains. Ambiguity was modulated by varying the quantity of hidden space on a bar (grey). (D) Proportion of risky choices according to the gamble win probability, against an ambiguous cue. Error bars indicate the standard error of the mean. Dotted lines indicate linear regression fits. (E) Proportion of risky choices according to the gamble win probability, against an ambiguous cue, plotted in function of the ambiguous bars' proportion of ambiguity. Error bars indicate the standard error of the mean. Dotted lines indicate linear regression fits. A and B are from Li et al., 2017. C, D, E are from Li et al., 2015

However, this 'experiential as ambiguous' argument is questionable for several reasons.

First, when presenting the 100% ambiguous option against all experienced options, subjects prefer the latter at a frequency proportional to its expected value (although the choice frequency does not perfectly match objective values) and the frequency curve monotonically increases with

the expected value. This result suggests (among others mentioned in chapter 5) that experiential values are remembered, even though they are discarded when comparing experienced options to described ones. Experiential options are linked to past information when the 100% ambiguous option does not. Besides, it has been shown that experience reduces ambiguity aversion (Güney and Newell, 2015). Also our design quite differs from classical ambiguity versus risk experiments (Fig. 6.3A/C). In previous studies, it is often difficult to variate simultaneously both the degree of ambiguity (uncertainty) and the expected value of the ambiguous option (which remains fixed most of the time; see Fig. 6.3A/C), resulting in measurements imprecision. Thus, in our experiment, the ambiguous option cannot be confused with experienced options, as the latter are not purely ambiguous and bears past outcome information.

Second, the pattern of *experiential neglect* can not be pure ambiguity aversion, as experienced options with negative expected-value, are most of the time preferred against symbolic options equally punishing. It is only when expected values are positive, that subject display opposite preferences, and are risk-seeking. In fact, a study found a fourfold pattern of ambiguity aversion, that shows differences in the loss and gain domain. In Kocher et al. (2018), ambiguity aversion is found for moderate likelihood gain and low likelihood loss prospects. Conversely, ambiguity seeking is found for low likelihood gain prospects and moderate likelihood loss prospects. This fourfold pattern matches the results in our task, where the tendency to choose experiential options in the loss domain (i.e. against symbolic options which probability is below 50%) is counterbalanced by a preference for symbolic options in the gain domain (i.e. against symbolic options which probability is above 50%). However, as this pattern isn't well substantiated yet.

Third, neural evidence suggests, that if risk and ambiguity are neurally represented in different ways (e.g. Hsu et al., 2005; Lauriola et al., 2007), ambiguity and RL subjective valuation might engage distinct circuitry as well (Bach et al., 2011).

Finally, if it is not a pure ambiguity aversion preference pattern, option learned through reinforcement likely carry ambiguous information. *Experiential neglect* could possibly be induced by the use of a heuristic in order to resolve this ambiguity, or more generally value-free decision processes. For instance, recently, [Pleskac and Hertwig \(2014\)](#) examined the relationship between probability and payoff in various environments, notably life insurance, dairy farming, or academic publishing. They highlighted that in a lot of natural environment, there is a legitimate belief for an inverse relation between payoffs and probabilities. For example, journals with a higher impact factor have a lower acceptance rate. They further tested this 'risk-reward' heuristic experimentally. They observed that the ambiguous or uncertain option became increasingly undesirable as the magnitude of the payoff increased. Indeed the 'risk-reward' heuristic predicts that high-reward options have low probability. This result may partially account for the low attractivity and underweighting of high-reward experiential options in our task.

Thus, the role of ambiguity aversion (as well as the possible heuristics behind it) and its relation to our behavioral pattern should be further investigated. Eventually, further imagery studies should disentangle the neural substrates involved in different forms of uncertainty (e.g. ambiguous options vs experiential options learned by reinforcement).

6.2.3 Fast-and-frugal heuristics

Does the discarding of experiential information observed in our subjects results from the use of a heuristic? Ignoring information, even relevant one, in order to maximize an effort-accuracy trade-off is a hallmark of heuristics as conceived by the fast-and-frugal research program ([Gigerenzer and Gaissmaier, 2011](#)).

Our subjects, on average, behave as-if they roughly follow a simple decision rule: choose the symbolic option if it has a positive expected value, otherwise choose the experiential option. To some extent, this decision process might belong to the class of heuristics that bases judgments on one good reason only, and ignore other information ([Gigerenzer and Gaissmaier, 2011](#)). Thus, a

one-clever-cue heuristic in our task, would consist in observing the proportion of red and green in the symbolic option, and choose the latter when green is dominant.

In Simon's scissors analogy:

Human rational behavior (and the rational behavior of all physical symbol systems) is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor (Simon, 1990).

According to the scissors analogy, this behavioral pattern is the consequence of our task structure coupled to limited computational resources. That said, is this behavior ecologically rational? Does it maximize an accuracy-effort trade-off? Apart from experiment 7, subjects perform above 80% of choices maximizing expected value in the Experiential-Symbolic condition (Fig. 6.4).

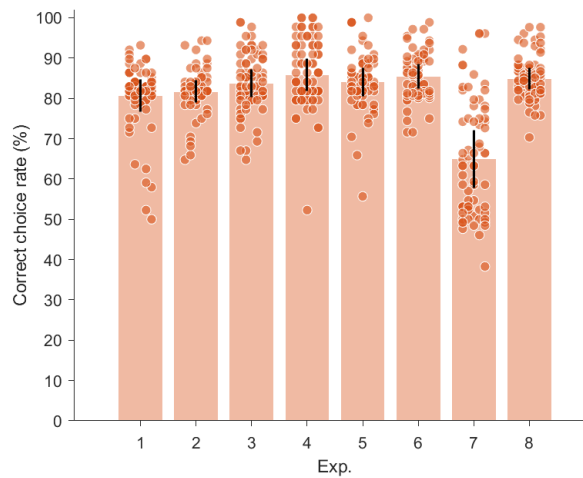


Figure 6.4: Performance from the Experiential-Symbolic phase described in chapter 5, among experiments 1-to-8. Each orange point represent a subject average performance. Error bars represent standard deviation.

In experiment 7, we changed the probabilities of experiential and symbolic options in order to test the *experience neglect* persistence in an environment where it would be ineffective. Interestingly, subjects kept discarding experiential information, resulting in a significant performance drop and economic loss.

As [Gigerenzer and Brighton \(2009\)](#) state it:

All inductive processes, including heuristics, make bets. This is why a heuristic is not inherently good or bad, or accurate or inaccurate, as is sometimes believed. Its accuracy is always relative to the structure of the environment. The study of the ecological rationality asks the following question: In which environments will a given heuristic succeed, and in which will it fail?

We identified a laboratory environment in which the *experiential neglect* pattern fails to provide good performance. Further research could consist in identifying natural environments where this pattern could be useful, as heuristics are supposed to result from evolutionary processes ([Gigerenzer, 2008](#)). In fact many decision situations involve a critical tension between experiential and symbolic values. Natural environments are filled with symbolic informations, and humans spontaneously build internal representations through experience ([Pitt, 2020](#)). For instance, when choosing a restaurant, one might feel conflicted between ratings provided by other users, and its personal experience.

Moreover, we can draw two interpretations by assuming that the pattern *experiential neglect* results from the use of a heuristic. Either the heuristic is mobilized in order to compare two types of values that are usually incommensurable. This implies that we are still engaged in an ontological claim about valuation (as we have been so far), since we assume that values are physically represented but perhaps in various forms. Alternatively, we could postulate that heuristics and value-free decision models are spontaneous modes of decision-making, and thus reject commitments about value representations.

6.2.4 Policy-based models and value as a reification

In a recent article, [Hayden and Niv \(2020\)](#) made the case for alternative approaches to the traditional two-step decision-making model, by addressing the *common currency hypothesis*. They have done so by challenging both ontological assumptions about subjective valuation and the epistemological responses that follow from them. Their arguments can roughly be broken down into

three claims.

First, neuroeconomics (Camerer et al., 2004b; Rustichini, 2009) inherited the utility measurement tradition. Since two decades, it proceeded to reify² the value construct by giving it a biological basis. Yet, there is mixed evidence for the brain to encode value *per se*, and *a fortiori* in a cardinal sense. The *common currency hypothesis* was established from the observation that firing rates of single neurons or neuronal populations correlate with values of outcomes (Kable and Glimcher, 2009; Rangel et al., 2008; Levy and Glimcher, 2012; O’Doherty, 2014), and that variations in these firing rates entails choice modulation (Sugrue et al., 2004; Strait et al., 2015). With the accumulation of neural evidence, it has been hypothesized that a modular valuation system guides decision-making. The OFC neural signals were supposed to encode value (experienced utility), when the PFC would drive choices³ (decision utility) (Kahneman et al., 1997; Kable and Glimcher, 2009; Levy and Glimcher, 2012; Bartra et al., 2013). In addition, the authors note that numerous findings in fact support a code for a relative preference of options (e.g. Tremblay and Schultz, 1999; Padoa-Schioppa, 2009; Klein et al., 2017), which seemingly excludes the idea of an abstract value code (i.e. cardinal utility) implemented in brain. Indeed, various evidence might corroborate this claim. For instance, when presenting one option at a time, neural responses do correlate with the value of the first presented option. However, the second (alternative) option correlate with the value difference, that is, a comparison process (Strait et al., 2015; Azab and Hayden, 2018). Moreover, they argue that skepticism toward relative (or ordinal) value coding is also legitimate, given that many neural findings could be reinterpreted as signals merely encoding outcome identity (Klein-Flügge et al., 2013; Rich and Wallis, 2016). In addition of outcome identity, they note that many confounds might prevent from identifying any “pure” neural correlates of value, such as the broad category of visceral, autonomic, skeletomotor processes (O’Doherty, 2014) or other properties (surpriseness, informativeness, etc.) of stimuli themselves (e.g. Wilson et al., 2014; Yoo and Hayden, 2018; Botvinik-Nezer et al., 2020)

²A reification is an abstraction treated as if it were a physical entity.

³This specialization of the PFC is however debated (Bartra et al., 2013)

Second, they build a philosophical argument, in which they argue that current methods are in fact unable to read out the “true subjective value” of an option. The only information we can get from choices is a rank relation between options, i.e. preferences. In fact, options values are inferred from preferences, but elicited preferences are noisy measurements that are sensible to confounding factors, as mentioned above. According to them, a final guardian knot that prevents from estimating “true” values, is thus the impossibility of obtaining value measurements independently of behavior. To make their point, they propose a thought experiment. Suppose a particular class of neurons whose firing rates are perfectly correlated with value inferred from preferences. In order to falsify the decision model, we manipulate the task. These manipulations induce changes of the firing rates that are perfectly consistent with our predictions. Hypothetically, these neurons could encode the value of our options. However, it has been shown that two preference sets can coexist under a single set of options (Schonberg and Katz, 2020). It entails that some neurons must encode option values, when others code for preferences. However, disentangling value neurons from preference neurons implies to show that preference neurons do not follow the assumed value function. However the value function is by construction inferred from preferences, resulting in a paradox.

To sum up, they argue that the current literature points toward a relative neural valuation (or even a pure comparison) of options, rather than an absolute cardinal assessment. Second, they claim that neural value, is in fact theoretically intractable, as it is inevitably tied to preferences as inferred from behavior.

Third, to overcome these problems, they propose to reject the epistemic constraint inherited from economic value theories, and endorse value-free approaches, such as policy-learning models (Bennett et al., 2021). Indeed, valuation is costly (Payne et al., 1992), and probably often not necessary. The success of heuristics (Gigerenzer and Gaissmaier, 2011), shows that valuation might not be the

default decision mode, even more so when items are fundamentally different. Computing action-values is one mean to maximize expected value (e.g. Q-learning; [Watkins, 1989](#)). Instead, policy models directly learn action policies. Policy-gradient models and Actor-Critic architectures belong to this class of models (see chapter 2 for a more detailed description). Besides, these models have found empirical support both in the behavioral and neuroscience literature (e.g. [Sakai and Fukai, 2008](#); [Maia, 2010](#); [O’Doherty et al., 2004](#); [Takahashi et al., 2008](#); [Colas et al., 2017](#)).

By actively avoiding invoking the theoretical construct of value, [Hayden and Niv \(2020\)](#) claim that their approach is more parsimonious, as it omits a supposedly unnecessary entity to explain decisions. Nonetheless, is value necessary to explain our *experiential neglect* pattern? As seen in previous sections, the pattern in itself might be partially explained by heuristics or ambiguity aversive attitudes. However, policy-learning might not be the best candidate to explain our results. During the learning phase, experiential options were presented in fixed pairs. Policy-learning predicts that individuals would learn the optimal policy for a given pair, while not being able to generalize to new decisions problems. Yet, experiential options are in average ranked correctly when presented (for the first time) against each others (Experiential-Experiential phase). Also, the subjective values inferred from these choices are in line with objective values. This result suggests that our subjects are able to extrapolate their experience within an option pair to new contexts, which requires at some point to project all the options into a new space that makes them comparable. So does the Stated Probabilities phase, in which subjective explicit ratings are also more rational than values elicited from experiential vs symbolic comparisons.

6.2.5 Are values built *a posteriori*?

We identified several levels at which experiential and symbolic values could diverge in construction. However, there may be a more fundamental epistemological problem with value. As said previously, value is inferred from preferences, which themselves are elicited from choices. There are situations in which, assuming a unique value function, different elicitation procedures elicit

different preferences. To explain this, [Lichtenstein and Slovic \(2006\)](#) hypothesized that preferences are generated at the time of elicitation rather than arising from any inherent value function. This violation of procedure invariance has been discussed by many scholars (e.g. [Tversky and Shafir, 1992](#), [Ariely et al., 2003](#)), often leading to the conclusion that talking about true preferences is irrelevant, in any normatively significant sense.

As [Hayden and Niv \(2020\)](#) state it:

That is, in the view of these and like-minded scholars, value doesn't sit in the brain waiting to be used; rather, preference is a complex and active process that takes place at the time the decision is made.

This may be reminiscent of the indeterminacy problems encountered in several disciplines, where measuring properties of a system's state possibly affects the state itself, and in turn the measurements. This is however not that surprising, as the brain and embodied mind are complex systems in perpetual motion ([Varela et al., 1992](#)). In a strong sense, it entails that preference elicitation is a circular process, as the measure creates the preference state. In a weaker (and probably more reasonable) sense, it suggests that preference elicitation does raise measurement uncertainty and falsification problems ([Glimcher, 2005](#)).

Taking this problem seriously might imply to foster decision models that build preferences on the fly, without assuming a unique set of preferences under a single value function ([Hayden and Niv, 2020](#); [Bennett et al., 2021](#)). However, the gain in flexibility could be at the expense of the ability to track regularities in decisions. These regularities probably emerge from neural and cognitive processes presenting a certain stability and permanence in time and space. Indeed, in our experiment, each elicitation (Experiential-Experiential, Experiential-Symbolic, Stated Probabilities) yields different subjective values, violating procedure invariance. Nevertheless, elicited values are well-ordered with regards to expected values, suggesting underlying stable valuation processes.

6.3 Conclusion

In the modern world, symbolic information and values are pervasive. From prices to weather forecasting, there exists a constant tension between these external information and our own internal representations. But how do we humans represent information internally? One answer would involve electrochemical reactions in the brain. However, this answer might be unsatisfying, as it does not tell much about cognition. Similarly, the french parliament is made of stones, it does not tell much about its social function and internal structure. In other words, we must go beyond the simple constituents. Rather, the debate on representations is focused on how information is represented and processed (Pitt, 2020). For instance, in the "imagery debate" (Pylyshyn and Dupoux, 2001), Marr (1982) proposed that visual representations are stored in a symbolic format (language-like, described primitives of objects) at an early stage of processing. Some opposed this thesis, yet no one disputed whether visual content is stored or not.

What about value representations? The picture might be less clear than in the case of vision. In chapter 1 and 2, we saw how two paradigms (experience and description) made the assumptions that individuals assign different scalar values to options, and afterwards select the highest. Marr's level has been applied to value based decision-making, and more generally in RL, as it translates well (Box 2.2). Undoubtedly, individuals are able to produce numerical estimates of average experienced outcomes, when they are asked to. In the same manner as they do when they draw an object from memory. It suggests that some "primitives" of value are stored during the learning process.

However, the level at which value is translated, and the format of its representation remains unclear. Indeed, debates on whether activity recorded in the OFC reflects value signals are raging (e.g. O'Doherty, 2014; Hayden and Niv, 2020). Regarding the format of value, as shown by context-dependence (Palminteri et al., 2015; Bavard et al., 2018, 2021), valuation might be performed in a

relative manner, and therefore relative values might be stored. In chapter 3 and 4, we discussed the *description-experience gap*, i.e. systematic behavioral discrepancies reported between the experience and description paradigms (Hertwig and Erev, 2009b; Wulff et al., 2018). We further hypothesized in chapter 5 that subjective representations of experienced and described values may systematically differ (Kellen et al., 2016). We concluded that the *description-experience gap* might also be instantiated at the individual level, and expressed by an *incommensurability* of symbolic and experiential values. In other words, symbolic and experiential values may be stored in different formats (i.e. *representational gap*).

Yet, to a certain extent, our *experiential neglect* pattern could actually be explained by value-free processes, notably heuristics. The fact that values can be inferred does not entail that all learning or decision processes involve value representations. Some decision problems might require value calculation, when others might not (e.g. see Appendix; Payne et al., 1992; Gigerenzer, 2008; Juechems and Summerfield, 2019; Hayden and Niv, 2020). Furthermore, regardless of whether a decision process involves value calculation, a single choice set can often be explained by different decision models (e.g. one value-based and one value-free). To some extent, this underdetermination⁴ (Duhem, 1991; Stanford, 2021) can be solved by rigorous falsification methods (Palminteri et al., 2017b). Nevertheless, a more precautionary position would be to favor pluralistic⁵ approaches to value based decision-making, in the tradition of cognitive science. For instance, although they are relatively incipient, connectionists⁶ alternatives have been proposed (Suri et al., 2020). This paradigm conceives value as an emergent phenomenon (and eventually suggests that value may simply not be represented; see Hunt and Hayden, 2017) and allows to relax assumptions such as the maximization of expected value.

Also, another obstacle to the idea of value representation stems from the polysemy surrounding

⁴The premise behind underdetermination of scientific theory is that the evidence available to us at any given time may be insufficient to identify what beliefs (here models) we should have in response to it.

⁵Or even perhaps, an "anything goes" approach (Feyerabend et al., 1993)

⁶Value based decision-making is often conceived within a modular architecture (Fodor, 1983). In contrast, connectionism models mental or behavioral phenomena as emergent processes of networks of simple interconnected units.

the concepts of both value (O’Doherty, 2014; Hayden and Niv, 2020) and representation (Poldrack, 2021). As Russ Poldrack (2021) states:

The ontological status and epistemic utility of mental representations are topics of enduring debate within the philosophy of mind. Neuroscientists have forged ahead largely unaware of these debates, using the term widely to describe the systematic empirical relationships that are often found to exist between neural activity and features of the external world.

A solution provided by Poldrack is to follow a set of criteria that demonstrate that these posited representations fulfill the “job description” for doing real representational work (Ramsey, 2007).

Last but not least, one should be careful about the legacy of previous theories. Value based decision-making inherits from a long tradition of theoretical and experimental work, where psychological and ontological commitments were often explicitly rejected. In contrast, neuroscience tends to naturalize and reify phenomena or abstract constructs, that are historically, socially or politically situated (Uttal, 2001; Choudhury et al., 2009; Joel and Fausto-Sterling, 2016; Poldrack, 2018; Hayden and Niv, 2020). Although there is no definitive solution to this risk, some answers might be found in the ‘critical neuroscience’ program⁷ (Choudhury et al., 2009).

⁷Inspired by the Frankfurt School (Max et al., 2017) it proposes a set of “self-critical practices, which aim to achieve reflective awareness of the standpoint-specific biases and constraints that enter into the production, interpretive framing and subsequent application of neuroscientific knowledge.” (Choudhury et al., 2009)

References

- M. Allais. Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica*, 21(4):503–546, 1953. ISSN 0012-9682. doi: 10.2307/1907921. URL <https://www.jstor.org/stable/1907921>. Publisher: [Wiley, Econometric Society].
- Dan Ariely, George Loewenstein, and Drazen Prelec. “coherent arbitrariness”: Stable demand curves without stable preferences. *The Quarterly journal of economics*, 118(1):73–106, 2003.
- Kenneth J Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, pages 265–290, 1954.
- Peter Ayton and Ilan Fischer. The hot hand fallacy and the gambler’s fallacy: Two faces of subjective randomness? *Memory & cognition*, 32(8):1369–1378, 2004.
- Habiba Azab and Benjamin Y Hayden. Correlates of economic decisions in the dorsal and subgenual anterior cingulate cortices. *European Journal of Neuroscience*, 47(8):979–993, 2018.
- Dominik R Bach, Oliver Hulme, William D Penny, and Raymond J Dolan. The known unknowns: neural representation of second-order uncertainty, and ambiguity. *Journal of Neuroscience*, 31(13):4811–4820, 2011.
- H Kent Baker and John R Nofsinger. *Behavioral finance: investors, corporations, and markets*, volume 6. John Wiley & Sons, 2010.
- Christian Balkenius, Jan Morén, et al. Computational models of classical conditioning: a comparative study. 1998.
- Moshe Bar. Visual objects in context. *Nature Reviews Neuroscience*, 5(8):617–629, 2004.
- Nicholas C Barberis. Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, 27(1):173–96, 2013.
- Jonathan Baron. *Rationality and intelligence*. Cambridge University Press, 1995.
- Jonathan Baron, D Koehler, and N Harvey. Normative models of judgment and decision making. *Blackwell handbook of judgment and decision making*, 2004.
- Greg Barron and Ido Erev. Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, 16(3):215–233, 2003a.

- Greg Barron and Ido Erev. Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, 16(3):215–233, 2003b. ISBN: 0894-3257 Publisher: Wiley Online Library.
- Oscar Bartra, Joseph T. McGuire, and Joseph W. Kable. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage*, 76:412–427, 2013. ISBN: 1053-8119 Publisher: Elsevier.
- Sophie Bavard. Mécanismes computationnels de l'apprentissage par renforcement dans les états sain et pathologique. 2021.
- Sophie Bavard, Maël Lebreton, Mehdi Khamassi, Giorgio Coricelli, and Stefano Palminteri. Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nature communications*, 9(1):1–12, 2018.
- Sophie Bavard, Aldo Rustichini, and Stefano Palminteri. Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Science Advances*, 7(14):eabe0340, 2021.
- Jonathan Baxter and Peter L Bartlett. Direct gradient-based reinforcement learning: I. gradient estimation algorithms. Technical report, Citeseer, 1999.
- Hannah M Bayer and Paul W Glimcher. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47(1):129–141, 2005.
- David E Bell, Howard Raiffa, Amos Tversky, et al. Descriptive, normative, and prescriptive interactions in decision making. *Decision making: Descriptive, normative, and prescriptive interactions*, 1:9–32, 1988.
- Richard Bellman. Dynamic programming and lagrange multipliers. *Proceedings of the National Academy of Sciences of the United States of America*, 42(10):767, 1956.
- Daniel Bennett, Yael Niv, and Angela J Langdon. Value-free reinforcement learning: policy optimization as a minimal model of operant behavior. *Current Opinion in Behavioral Sciences*, 41: 114–121, 2021.
- Jeremy Bentham. *The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation*. Clarendon Press, 1789.
- Nathan Berg, Guido Biele, and Gerd Gigerenzer. Does consistency predict accuracy of beliefs?: Economists surveyed about psa. *Economists Surveyed About PSA*, 2010.
- Henri Bergson. Time and free will, new york (harper & row) 1960. 1889.
- Daniel Bernoulli. Exposition of a new theory on the measurement of risk. In *The Kelly capital growth investment criterion: Theory and practice*, pages 11–24. World Scientific, 2011.
- Alfred Binet and Th Simon. Le développement de l'intelligence chez les enfants. *L'année psychologique*, 14(1):1–94, 1907.

References

- Tapan Biswas. The expected utility theory. In *Decision-Making under Uncertainty*, pages 3–18. Springer, 1997.
- Han Bleichrodt and Peter P Wakker. Regret theory: A bold alternative to the alternatives. *The Economic Journal*, 125(583):493–532, 2015.
- Lawrence A Boland. A critique of friedman’s critics. *Journal of Economic literature*, 17(2):503–522, 1979.
- Rotem Botvinik-Nezer, Felix Holzmeister, Colin F Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A Mumford, R Alison Adcock, et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582(7810): 84–88, 2020.
- Eduard Brandstätter, Gerd Gigerenzer, and Ralph Hertwig. The priority heuristic: making choices without trade-offs. *Psychological review*, 113(2):409, 2006.
- Arndt Bröder and Stefanie Schiffer. Take the best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132(2):277, 2003.
- Joshua W Brown and Todd S Braver. Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, 307(5712):1118–1121, 2005.
- Joshua W Brown and Todd S Braver. A computational model of risk, conflict, and individual difference effects in the anterior cingulate cortex. *Brain research*, 1202:99–108, 2008.
- Colin Camerer and Martin Weber. Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of risk and uncertainty*, 5(4):325–370, 1992.
- Colin Camerer, George Loewenstein, and Drazen Prelec. Neuroeconomics: How neuroscience can inform economics. *Journal of economic Literature*, 43(1):9–64, 2005.
- Colin F Camerer, George Loewenstein, and Drazen Prelec. Neuroeconomics: Why economics needs brains. *scandinavian Journal of Economics*, 106(3):555–579, 2004a.
- Colin F. Camerer, George Loewenstein, and Drazen Prelec. Neuroeconomics: Why economics needs brains. *scandinavian Journal of Economics*, 106(3):555–579, 2004b. ISBN: 0347-0520 Publisher: Wiley Online Library.
- Anjan Chakravartty. Scientific Realism. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2017 edition, 2017.
- M Keith Chen, Venkat Lakshminarayanan, and Laurie R Santos. How basic are behavioral biases? evidence from capuchin monkey trading behavior. *Journal of political economy*, 114(3):517–537, 2006.
- Vikram S Chib, Antonio Rangel, Shinsuke Shimojo, and John P O’Doherty. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, 29(39):12315–12320, 2009.

- Suparna Choudhury, Saskia Kathi Nagel, and Jan Slaby. Critical neuroscience: Linking neuroscience and society through critical practice. *BioSocieties*, 4(1):61–77, 2009.
- Yun-Peng Chu and Ruey-Ling Chu. The subsidence of preference reversals in simplified and marketlike experimental settings: A note. *The American Economic Review*, 80(4):902–911, 1990.
- Jaron T Colas, Wolfgang M Pauli, Tobias Larsen, J Michael Tyszka, and John P O’Doherty. Distinct prediction errors in mesostriatal circuits of the human brain mediate learning about the values of both states and actions: evidence from high-resolution fmri. *PLoS computational biology*, 13(10):e1005810, 2017.
- Howard C Cromwell, Oum K Hassani, and Wolfram Schultz. Relative reward processing in primate striatum. *Experimental Brain Research*, 162(4):520–525, 2005.
- Shawn P Curley, J Frank Yates, and Richard A Abrams. Psychological sources of ambiguity avoidance. *Organizational behavior and human decision processes*, 38(2):230–256, 1986.
- Jean Czerlinski, Gerd Gigerenzer, and Daniel G Goldstein. How good are simple heuristics? In *Simple heuristics that make us smart*, pages 97–118. Oxford University Press, 1999.
- Nathaniel D. Daw and Kenji Doya. The computational neurobiology of learning and reward. *Current opinion in neurobiology*, 16(2):199–204, 2006. ISBN: 0959-4388 Publisher: Elsevier.
- Nathaniel D Daw, John P O’doherly, Peter Dayan, Ben Seymour, and Raymond J Dolan. Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–879, 2006.
- Benedetto De Martino, Dharshan Kumaran, Ben Seymour, and Raymond J. Dolan. Frames, biases, and rational decision-making in the human brain. *Science*, 313(5787):684–687, 2006. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science.
- Francesca De Petrillo, Marialba Ventricelli, Giorgia Ponsi, and Elsa Addessi. Do tufted capuchin monkeys play the odds? flexible risk preferences in sapajus spp. *Animal cognition*, 18(1):119–130, 2015.
- Mauricio R Delgado, Melinda M Miller, Souheil Inati, and Elizabeth A Phelps. An fmri study of reward-related probability learning. *Neuroimage*, 24(3):862–873, 2005.
- Mauricio R Delgado, Rita L Jou, and Elizabeth A Phelps. Neural systems underlying aversive conditioning in humans with primary and secondary reinforcers. *Frontiers in neuroscience*, 5:71, 2011.
- Anja Dieckmann and Jörg Rieskamp. The influence of information redundancy on probabilistic inferences. *Memory & Cognition*, 35(7):1801–1813, 2007.
- Dudley Dillard. The status of the labor theory of value. *Southern Economic Journal*, pages 345–352, 1945.
- Pierre Maurice Marie Duhem. *The aim and structure of physical theory*, volume 13. Princeton University Press, 1991.

References

- Tsvetomira Dumbalska, Vickie Li, Konstantinos Tsetos, and Christopher Summerfield. A map of decoy influence in human multialternative choice. *Proceedings of the National Academy of Sciences*, 117(40):25169–25178, 2020.
- Hermann Ebbinghaus. Memory: A contribution to experimental psychology. *Annals of neurosciences*, 20(4):155, 1913.
- Francis Ysidro Edgeworth. *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. Number 10. CK Paul, 1881.
- Ward Edwards. Conservatism in human information processing. *Formal representation of human judgment*, 1968.
- Hillel J Einhorn and Robin M Hogarth. Ambiguity and uncertainty in probabilistic inference. *Psychological review*, 92(4):433, 1985.
- Daniel Ellsberg. Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, 75(4):643–669, 1961. ISSN 0033-5533. doi: 10.2307/1884324. URL <https://www.jstor.org/stable/1884324>.
- Robert Epstein. The principle of parsimony and some applications in psychology. *The Journal of Mind and Behavior*, pages 119–130, 1984.
- Ido Erev and Alvin E Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, pages 848–881, 1998.
- Ido Erev, Ira Glozman, and Ralph Hertwig. What impacts the impact of rare events. *Journal of Risk and Uncertainty*, 36(2):153–177, 2008.
- Ido Erev, Eyal Ert, Alvin E Roth, Ernan Haruvy, Stefan M Herzog, Robin Hau, Ralph Hertwig, Terrence Stewart, Robert West, and Christian Lebiere. A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, 23(1):15–47, 2010.
- Ido Erev, Eyal Ert, Ori Plonsky, Doron Cohen, and Oded Cohen. From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychological review*, 124(4):369, 2017. ISBN: 1939-1471 Publisher: American Psychological Association.
- Gustav Theodor Fechner. *Elements of psychophysics*, 1860. 1860.
- Paul Feyerabend et al. *Against method*. Verso, 1993.
- Christopher D Fiorillo, Philippe N Tobler, and Wolfram Schultz. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614):1898–1902, 2003.
- Peter C Fishburn. *Utility theory for decision making*. Technical report, Research analysis corp McLean VA, 1970.
- Irving Fisher and William Joseph Barber. *The rate of interest*. Garland Pub., 1907.

- Thomas H. B. FitzGerald, Ben Seymour, Dominik R. Bach, and Raymond J. Dolan. Differentiable Neural Substrates for Learned and Described Value and Risk. *Current Biology*, 20(20):1823–1829, October 2010. ISSN 0960-9822. doi: 10.1016/j.cub.2010.08.048. URL <http://www.sciencedirect.com/science/article/pii/S0960982210010705>.
- Jerry A Fodor. *The modularity of mind*. MIT press, 1983.
- Craig R Fox and Amos Tversky. Ambiguity aversion and comparative ignorance. *The quarterly journal of economics*, 110(3):585–603, 1995.
- Craig R Fox and Amos Tversky. A belief-based account of decision under uncertainty. *Management science*, 44(7):879–895, 1998.
- Craig R Fox and Martin Weber. Ambiguity aversion, comparative ignorance, and decision context. *Organizational behavior and human decision processes*, 88(1):476–498, 2002.
- Anthony NS Freeling. A philosophical basis for decision aiding. *Theory and Decision*, 16(2):179–206, 1984.
- Milton Friedman. *Essays in Positive Economics*. University of Chicago Press, 1953. ISBN 978-0-226-26403-5. Google-Books-ID: Fv8846OSbvWC.
- Milton Friedman and Leonard J Savage. The utility analysis of choices involving risk. *Journal of political Economy*, 56(4):279–304, 1948.
- Deborah Frisch and Jonathan Baron. Ambiguity and rationality. *Journal of Behavioral Decision Making*, 1(3):149–157, 1988.
- EA Gaffan. Review: An introduction to animal cognition, 1989.
- David Genesove and Christopher Mayer. Loss aversion and seller behavior: Evidence from the housing market. *The quarterly journal of economics*, 116(4):1233–1260, 2001.
- Gerd Gigerenzer. How to make cognitive illusions disappear: Beyond “heuristics and biases”. *European review of social psychology*, 2(1):83–115, 1991.
- Gerd Gigerenzer. On narrow norms and vague heuristics: A reply to Kahneman and Tversky. 1996.
- Gerd Gigerenzer. Why heuristics work. *Perspectives on psychological science*, 3(1):20–29, 2008.
- Gerd Gigerenzer. The bias bias in behavioral economics. *Review of Behavioral Economics*, 5(3-4):303–336, 2018.
- Gerd Gigerenzer and Henry Brighton. Homo heuristicus: Why biased minds make better inferences. *Topics in cognitive science*, 1(1):107–143, 2009.
- Gerd Gigerenzer and Wolfgang Gaissmaier. Heuristic decision making. *Annual review of psychology*, 62:451–482, 2011.

References

- Gerd Gigerenzer and Daniel G Goldstein. Betting on one good reason: The take the best heuristic. In *Simple heuristics that make us smart*, pages 75–95. Oxford University Press, 1999.
- Gerd Gigerenzer, Ralph Hertwig, Eva Van Den Broek, Barbara Fasolo, and Konstantinos V Katsikopoulos. “a 30% chance of rain tomorrow”: How does the public understand probabilistic weather forecasts? *Risk Analysis: An International Journal*, 25(3):623–629, 2005.
- Thomas Gilovich, Robert Vallone, and Amos Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology*, 17(3):295–314, 1985.
- Nicola Giocoli. Modeling rational agents the consistency view of rationality and the changing image of neoclassical economics:. *Cahiers d'économie Politique*, n° 49(2):177–208, December 2005. ISSN 0154-8344. doi: 10.3917/cep.049.0177. URL <https://www.cairn.info/revue-cahiers-d-economie-politique-1-2005-2-page-177.htm?ref=doi>.
- Paul W Glimcher. Indeterminacy in brain and behavior. *Annu. Rev. Psychol.*, 56:25–56, 2005.
- Gary H Glover. Overview of functional magnetic resonance imaging. *Neurosurgery Clinics*, 22(2):133–139, 2011.
- William M. Goldstein and Robin M. Hogarth. Judgment and decision research: Some historical context. 1997. ISBN: 0521483026 Publisher: Cambridge University Press.
- George Graham. Behaviorism. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2019 edition, 2019.
- Faruk Gul and Wolfgang Pesendorfer. The case for mindless economics. *The foundations of positive and normative economics: A handbook*, 1:3–42, 2008.
- Şule Güney and Ben R Newell. Overcoming ambiguity aversion through experience. *Journal of Behavioral Decision Making*, 28(2):188–199, 2015.
- Norman Guttman. On skinner and hull: A reminiscence and projection. *American Psychologist*, 32(5):321, 1977.
- Ulrike Hahn and Paul A Warren. Perceptions of randomness: Why three heads are better than four. *Psychological review*, 116(2):454, 2009.
- Yoram Halevy. Ellsberg revisited: An experimental study. *Econometrica*, 75(2):503–536, 2007.
- Todd A Hare, John O’doherly, Colin F Camerer, Wolfram Schultz, and Antonio Rangel. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *Journal of neuroscience*, 28(22):5623–5630, 2008.
- David W Harless and Colin F Camerer. The predictive utility of generalized expected utility theories. *Econometrica: Journal of the Econometric Society*, pages 1251–1289, 1994.
- Benjamin Hayden and Yael Niv. The case against economic values in the brain. October 2020. doi: 10.31234/osf.io/7hgup. URL <https://psyarxiv.com/7hgup/>. Publisher: PsyArXiv.

- Benjamin Hayden, Sarah Heilbronner, and Michael Platt. Ambiguity aversion in rhesus macaques. *Frontiers in neuroscience*, 4:166, 2010. ISBN: 1662-453X Publisher: Frontiers.
- Benjamin Y Hayden and Yael Niv. The case against economic values in the orbitofrontal cortex (or anywhere else in the brain). *Behavioral Neuroscience*, 135(2):192, 2021.
- Benjamin Y Hayden and Michael L Platt. The mean, the median, and the st. petersburg paradox. *Judgment and Decision Making*, 4(4):256, 2009.
- Benjamin Y Hayden, Sarah R Heilbronner, John M Pearson, and Michael L Platt. Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *Journal of Neuroscience*, 31(11):4178–4187, 2011.
- Sarah Heilbronner, Benjamin Yost Hayden, and Michael Platt. Decision salience signals in posterior cingulate cortex. *Frontiers in neuroscience*, 5:55, 2011.
- Sarah R. Heilbronner and Benjamin Y. Hayden. Contextual Factors Explain Risk-Seeking Preferences in Rhesus Monkeys. *Frontiers in Neuroscience*, 7, 2013. ISSN 1662-4548. doi: 10.3389/fnins.2013.00007. URL <http://journal.frontiersin.org/article/10.3389/fnins.2013.00007/abstract>.
- Sarah R Heilbronner and Benjamin Y Hayden. The description-experience gap in risky choice in nonhuman primates. *Psychonomic bulletin & review*, 23(2):593–600, 2016.
- Richard J Herrnstein. Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the experimental analysis of behavior*, 4(3):267, 1961.
- Ralph Hertwig and Ido Erev. The description–experience gap in risky choice. *Trends in cognitive sciences*, 13(12):517–523, 2009a.
- Ralph Hertwig and Ido Erev. The description–experience gap in risky choice. *Trends in cognitive sciences*, 13(12):517–523, 2009b. ISBN: 1364-6613 Publisher: Elsevier.
- Ralph Hertwig and Timothy J Pleskac. The game of life: How small samples render choice simpler. *The probabilistic mind: Prospects for Bayesian cognitive science*, pages 209–235, 2008.
- Ralph Hertwig, Greg Barron, Elke U. Weber, and Ido Erev. Decisions from experience and the effect of rare events in risky choice. *Psychological science*, 15(8):534–539, 2004. ISBN: 0956-7976 Publisher: SAGE Publications Sage CA: Los Angeles, CA.
- Ralph Hertwig, Björn Benz, and Stefan Krauss. The conjunction fallacy and the many meanings of and. *Cognition*, 108(3):740–753, 2008.
- John R. Hicks. The theory of uncertainty and profit. *Economica*, (32):170–189, 1931. Publisher: JSTOR.
- Charles A Holt and Susan K Laury. Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655, 2002.
- Harold Hotelling. Stability in competition. *The Economic Journal*, 39(153):41–57, 1929.

References

- Ronald A Howard. Dynamic programming and markov processes. 1960.
- Ming Hsu, Meghana Bhatt, Ralph Adolphs, Daniel Tranel, and Colin F Camerer. Neural systems responding to degrees of uncertainty in human decision-making. *Science*, 310(5754):1680–1683, 2005.
- Scott A Huettel, Allen W Song, Gregory McCarthy, et al. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, MA, 2004.
- Clark L Hull. The goal-gradient hypothesis and maze learning. *Psychological review*, 39(1):25, 1932.
- Laurence T Hunt and Benjamin Y Hayden. A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews Neuroscience*, 18(3):172–182, 2017.
- Masaki Isoda. Socially relative reward valuation in the primate brain. *Current Opinion in Neurobiology*, 68:15–22, 2021.
- Katsuhito Iwai. Evolution of money. *Evolution of Economic Diversity*, pages 396–431, 1997.
- William Jaffé. Menger, jevons and walras de-homogenized. *Economic Inquiry*, 14(4):511–524, 1976.
- Ryan K. Jessup, Anthony J. Bishara, and Jerome R. Busemeyer. Feedback produces divergence from prospect theory in descriptive choice. *Psychological Science*, 19(10):1015–1022, 2008. ISBN: 0956-7976 Publisher: SAGE Publications Sage CA: Los Angeles, CA.
- W. Jevons. *The Theory of Political Economy*. Palgrave Classics in Economics. Palgrave Macmillan UK, 1871. ISBN 978-1-137-37414-1. doi: 10.1057/9781137374158. URL <https://www.palgrave.com/gp/book/9781137374141>.
- W. Jevons. *The Theory of Political Economy*. Palgrave Classics in Economics. Palgrave Macmillan UK, 2013. ISBN 978-1-137-37414-1. doi: 10.1057/9781137374158. URL <https://www.palgrave.com/gp/book/9781137374141>.
- Daphna Joel and Anne Fausto-Sterling. Beyond sex differences: new approaches for thinking about variation in brain structure and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1688):20150451, 2016.
- Keno Juechems and Christopher Summerfield. Where does value come from? *Trends in cognitive sciences*, 23(10):836–850, 2019.
- Joseph W Kable and Paul W Glimcher. The neurobiology of decision: consensus and controversy. *Neuron*, 63(6):733–745, 2009.
- Daniel Kahneman and Shane Frederick. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49:81, 2002.
- Daniel Kahneman and Amos Tversky. Subjective probability: A judgment of representativeness. *Cognitive psychology*, 3(3):430–454, 1972.

- Daniel Kahneman and Amos Tversky. On the psychology of prediction. *Psychological review*, 80 (4):237, 1973.
- Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):363–391, 1979.
- Daniel Kahneman and Amos Tversky. The simulation heuristic. Technical report, Stanford Univ CA Dept of Psychology, 1981.
- Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge university press, 1982a.
- Daniel Kahneman, Stewart Paul Slovic, Paul Slovic, and Amos Tversky. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, April 1982b. ISBN 978-0-521-28414-1. Google-Books-ID: _0H8gwj4a1MC.
- Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, 5(1):193–206, March 1991. ISSN 0895-3309. doi: 10.1257/jep.5.1.193. URL <https://www.aeaweb.org/articles?id=10.1257/jep.5.1.193>.
- Daniel Kahneman, Peter P Wakker, and Rakesh Sarin. Back to bentham? explorations of experienced utility. *The quarterly journal of economics*, 112(2):375–406, 1997.
- Leon J Kamin. Attention-like processes in classical conditioning. In *SYMP. ON AVERSIVE MOTIVATION MIAMI*, 1967a.
- Leon J Kamin. Predictability, surprise, attention, and conditioning. In *SYMP. ON PUNISHMENT*, 1967b.
- David Kellen, Thorsten Pachur, and Ralph Hertwig. How (in) variant are subjective representations of described and experienced risk and rewards? *Cognition*, 157:126–138, 2016.
- John Maynard Keynes. *The General Theory of Employment, Interest, and Money*. Springer, 1936. ISBN 978-3-319-70344-2. Google-Books-ID: Su1lDwAAQBAJ.
- J. E. King and Michael McLure. History of the Concept of Value. Technical Report 14-06, The University of Western Australia, Department of Economics, 2014. URL <https://ideas.repec.org/p/uwa/wpaper/14-06.html>. Publication Title: Economics Discussion / Working Papers.
- Tilman A Klein, Markus Ullsperger, and Gerhard Jocham. Learning relative values in the striatum induces violations of normative decision making. *Nature communications*, 8(1):1–12, 2017.
- Miriam Cornelia Klein-Flügge, Helen Catharine Barron, Kay Henning Brodersen, Raymond J Dolan, and Timothy Edward John Behrens. Segregated encoding of reward–identity and stimulus–reward associations in human orbitofrontal cortex. *Journal of Neuroscience*, 33(7):3202–3211, 2013.

References

- Frank Hyneman Knight. *Risk, uncertainty and profit*, volume 31. Houghton Mifflin, 1921.
- Brian Knutson, Jonathan Taylor, Matthew Kaufman, Richard Peterson, and Gary Glover. Distributed neural representation of expected value. *Journal of Neuroscience*, 25(19):4806–4812, 2005.
- Martin G Kocher, Amrei Marie Lahno, and Stefan T Trautmann. Ambiguity aversion is not universal. *European Economic Review*, 101:268–283, 2018.
- Christopher Krupenye, Alexandra G Rosati, and Brian Hare. Bonobos and chimpanzees exhibit human-like framing effects. *Biology letters*, 11(2):20140527, 2015.
- Anton Kühberger and Josef Perner. The role of competition and knowledge in the ellberg task. *Journal of Behavioral Decision Making*, 16(3):181–191, 2003.
- Thomas Kuhn. *The structure of scientific revolutions*. Princeton University Press, 2021.
- Kenneth K Kwong, John W Belliveau, David A Chesler, Inna E Goldberg, Robert M Weisskoff, Brigitte P Poncelet, David N Kennedy, Bernice E Hoppel, Mark S Cohen, and Robert Turner. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12):5675–5679, 1992.
- Peter Landry and Ryan Webb. Pairwise normalization: A neuroeconomic theory of multi-attribute choice. *Journal of Economic Theory*, 193:105221, 2021.
- Marco Lauriola, Irwin P Levin, and Stephanie S Hart. Common and distinct factors in decision making under ambiguity and risk: A psychometric study of individual differences. *Organizational Behavior and Human Decision Processes*, 104(2):130–149, 2007.
- Marvin L Leathers and Carl R Olson. In monkeys making value-based decisions, lip neurons encode cue salience and not action value. *Science*, 338(6103):132–135, 2012.
- Germain Lefebvre, Maël Lebreton, Florent Meyniel, Sacha Bourgeois-Gironde, and Stefano Palminteri. Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, 1(4):1–9, 2017. ISBN: 2397-3374 Publisher: Nature Publishing Group.
- Tomás Lejarraga and Johannes Müller-Trede. When experience meets description: How dyads integrate experiential and descriptive information in risky decisions. *Management Science*, 63(6):1953–1971, 2017.
- Thomas C Leonard. Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness, 2008.
- Dino J Levy and Paul W Glimcher. The root of all value: a neural common currency for choice. *Current opinion in neurobiology*, 22(6):1027–1038, December 2012. ISSN 0959-4388. doi: 10.1016/j.conb.2012.06.001. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4093837/>.
- Jack S Levy. Loss aversion, framing, and bargaining: The implications of prospect theory for international conflict. *International Political Science Review*, 17(2):179–195, 1996.

- Rosa Li, Elizabeth M Brannon, and Scott A Huettel. Children do not exhibit ambiguity aversion despite intact familiarity bias. *Frontiers in psychology*, 5:1519, 2015.
- Rosa Li, Rachel C Roberts, Scott A Huettel, and Elizabeth M Brannon. Five-year-olds do not show ambiguity aversion in a risk and ambiguity task with physical objects. *Journal of experimental child psychology*, 159:319–326, 2017.
- Sarah Lichtenstein and Paul Slovic. *The construction of preference*. Cambridge University Press, 2006. ISBN 1-139-45778-0.
- Graham Loomes and Robert Sugden. Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal*, 92(368):805–824, 1982. ISBN: 0013-0133 Publisher: JSTOR.
- Graham Loomes, Chris Starmer, and Robert Sugden. Are preferences monotonic? testing some predictions of regret theory. *Economica*, pages 17–33, 1992.
- Kenway Louie and Benedetto De Martino. The neurobiology of context-dependent valuation and choice. In *Neuroeconomics*, pages 455–476. Elsevier, 2014.
- Marsha C Lovett, Lynne M Reder, and Christian Lebiere. Modeling working memory in a unified architecture: An act-r perspective. 1999.
- R Duncan Luce. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.
- Elliot A Ludvig, Marc G Bellemare, and Keir G Pearson. A primer on reinforcement learning in the brain: Psychological, computational, and neural perspectives. *Computational neuroscience for advancing artificial intelligence: Models, methods and applications*, pages 111–144, 2011.
- Elliot A Ludvig, Christopher R Madan, Jeffrey M Pisklak, and Marcia L Spetch. Reward context determines risky choice in pigeons and humans. *Biology Letters*, 10(8):20140451, 2014.
- Christopher R Madan, Elliot A Ludvig, and Marcia L Spetch. Comparative inspiration: From puzzles with pigeons to novel discoveries with humans in risky choice. *Behavioural processes*, 160:10–19, 2019.
- Tiago V Maia. Two-factor theory, the actor-critic model, and conditioned avoidance. *Learning & behavior*, 38(1):50–67, 2010.
- Mauro Maldonato, Silvia Dell’Orco, et al. Decision making styles and adaptive algorithms for human action. *Psychology*, 2(08):811, 2011.
- John C Malone. Did john b. watson really “found” behaviorism? *The Behavior Analyst*, 37(1): 1–12, 2014.
- D. Marr. Vision: A computational investigation into the human representation and processing of visual information. 1982.
- David Marr and Tomaso Poggio. From understanding computation to understanding neural circuitry. 1976.

References

- Karl Marx. *Capital: A critical analysis of capitalist production*. Humboldt, 1873.
- John HR Maunsell. Neuronal representations of cognitive state: reward or attention? *Trends in cognitive sciences*, 8(6):261–265, 2004.
- Horkheimer Max, W Adorno Theodor, Torr Zoltán, and Landmann Michael. *The Frankfurt School*. Routledge, 2017.
- Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- Craig RM McKenzie. Judgment and decision making. *Handbook of cognition*, 2005.
- C Menger. *Principes d'économie (grundsätze der volkswirtschaftslehre)*, 1871.
- John Stuart Mill. Utilitarianism (1863). *Utilitarianism, Liberty, Representative Government*, pages 7–9, 1859.
- Kevin J Miller, Amitai Shenhav, and Elliot A Ludvig. Habits without values. *Psychological review*, 126(2):292, 2019.
- Ralph R Miller, Robert C Barnet, and Nicholas J Grahame. Assessment of the rescorla-wagner model. *Psychological bulletin*, 117(3):363, 1995.
- Philippe Mongin. The allais paradox: what it became, what it really was, what it now suggests to us. *Economics & Philosophy*, 35(3):423–459, 2019.
- J. W Moore and N. A. Schmajuk. Kamin blocking. *Scholarpedia*, 3(5):3542, 2008. doi: 10.4249/scholarpedia.3542. revision #89027.
- Mary S Morgan. Economic man as model man: ideal types, idealization and caricatures. *Journal of the History of Economic Thought*, 28(1):1–27, 2006.
- Genela Morris, David Arkadir, Alon Nevet, Eilon Vaadia, and Hagai Bergman. Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron*, 43(1):133–143, 2004.
- Ivan Moscati. *Measuring utility: From the marginal revolution to behavioral economics*. Oxford Studies in History of E, 2018a.
- Ivan Moscati. *Measuring Utility: From the Marginal Revolution to Behavioral Economics*. Oxford Studies in History of Economics. Oxford University Press, New York, 2018b. ISBN 978-0-19-937276-8. doi: 10.1093/oso/9780199372768.001.0001. URL <https://oxford.universitypressscholarship.com/10.1093/oso/9780199372768.001.0001/oso-9780199372768>.
- Ivan Moscati. Not a behaviorist: Samuelson's contributions to utility theory in the harvard years, 1936–1940. In *Paul Samuelson*, pages 243–278. Springer, 2019.

- Ivan Moscati et al. How cardinal utility entered economic analysis during the ordinal revolution. *Centro di Studi Sulla Storia ei Metodi dell'Economia Politica "Claudio Napoleoni", Working Paper*, (01):1–31, 2013.
- Anis Najar, Emmanuelle Bonnet, Bahador Bahrami, and Stefano Palminteri. The actions of others act as a pseudo-reward to drive imitation in the context of social reinforcement learning. *PLoS biology*, 18(12):e3001028, 2020.
- Sander Nieuwenhuis, Dirk J Heslenfeld, Niels J Alting von Geusau, Rogier B Mars, Clay B Holroyd, and Nick Yeung. Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage*, 25(4):1302–1309, 2005.
- Yael Niv. The primacy of behavioral research for understanding the brain. *Behavioral Neuroscience*, 2021.
- Yael Niv and Angela Langdon. Reinforcement learning with marr. *Current opinion in behavioral sciences*, 11:67–73, 2016.
- Yael Niv and Geoffrey Schoenbaum. Dialogues on prediction errors. *Trends in cognitive sciences*, 12(7):265–272, 2008.
- FJ Odling-Smee. Background stimuli and the inter-stimulus interval during pavlovian conditioning. *Quarterly Journal of Experimental Psychology*, 27(3):387–392, 1975.
- John O’Doherty, Peter Dayan, Johannes Schultz, Ralf Deichmann, Karl Friston, and Raymond J Dolan. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *science*, 304(5669):452–454, 2004.
- John P O’Doherty. The problem with value. *Neuroscience & Biobehavioral Reviews*, 43:259–268, 2014.
- Camillo Padoa-Schioppa. Range-adapting representation of economic value in the orbitofrontal cortex. *Journal of Neuroscience*, 29(44):14004–14014, 2009.
- Camillo Padoa-Schioppa and John A Assad. Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441(7090):223–226, 2006.
- Camillo Padoa-Schioppa and Katherine E Conen. Orbitofrontal cortex: a neural circuit for economic decisions. *Neuron*, 96(4):736–754, 2017.
- Stefano Palminteri and Maël Lebreton. Context-dependent outcome encoding in human reinforcement learning. 2021.
- Stefano Palminteri, Mehdi Khamassi, Mateus Joffily, and Giorgio Coricelli. Contextual modulation of value signals in reward and punishment learning. *Nature communications*, 6(1):1–14, 2015. ISBN: 2041-1723 Publisher: Nature Publishing Group.

References

- Stefano Palminteri, Germain Lefebvre, Emma J. Kilford, and Sarah-Jayne Blakemore. Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLoS computational biology*, 13(8):e1005684, 2017a. ISBN: 1553-7358 Publisher: Public Library of Science.
- Stefano Palminteri, Valentin Wyart, and Etienne Koechlin. The importance of falsification in computational cognitive modeling. *Trends in cognitive sciences*, 21(6):425–433, 2017b.
- Vilfredo Pareto. The new theories of economics. *Journal of political economy*, 5(4):485–502, 1897.
- Martin P Paulus and Lawrence R Frank. Anterior cingulate activity modulates nonlinear decision weight function of uncertain prospects. *Neuroimage*, 30(2):668–677, 2006.
- Ivan Petrovitch Pavlov and William Gantt. Lectures on conditioned reflexes: Twenty-five years of objective study of the higher nervous activity (behaviour) of animals. 1928.
- John W Payne, James R Bettman, Eloise Coupey, and Eric J Johnson. A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica*, 80(1-3): 107–141, 1992.
- Joseph Persky. The ethology of homo economicus. *Journal of Economic Perspectives*, 9(2):221–231, 1995.
- Mathias Pessiglione, Ben Seymour, Guillaume Flandin, Raymond J. Dolan, and Chris D. Frith. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442(7106):1042–1045, 2006. ISBN: 1476-4687 Publisher: Nature Publishing Group.
- David Pitt. Mental Representation. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2020 edition, 2020.
- Michael L Platt and Paul W Glimcher. Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741):233–238, 1999.
- Timothy J Pleskac and Ralph Hertwig. Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, 143(5):2000, 2014.
- Henri Poincaré. Le continu mathématique. *Revue de métaphysique et de morale*, 1(1):26–34, 1893.
- Russell A Poldrack. *The new mind readers: What neuroimaging can and cannot reveal about our thoughts*. Princeton University Press, 2018.
- Russell A Poldrack. The physics of representation. *Synthese*, 199(1):1307–1325, 2021.
- Andrea Polonioli. Re-assessing the heuristics debate, 2013.
- George Polya. *How to solve it: A new aspect of mathematical method*. Number 246. Princeton university press, 1945.
- Kerstin Preuschoff, Peter Bossaerts, and Steven R Quartz. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron*, 51(3):381–390, 2006.

Zenon Pylyshyn and E Dupoux. Is the imagery debate over? if so, what was it about. *Language, brain, and cognitive development: essays in honor of Jacques Mehler*. MIT Press, Cambridge, MA, pages 59–83, 2001.

William M Ramsey. *Representation reconsidered*. Cambridge University Press, 2007.

Antonio Rangel, Colin Camerer, and P Read Montague. A framework for studying the neurobiology of value-based decision making. *Nature reviews neuroscience*, 9(7):545–556, 2008.

Michel Regenwetter and Maria M Robinson. The construct–behavior gap in behavioral decision research: A challenge beyond replicability. *Psychological Review*, 124(5):533, 2017.

Robert A Rescorla. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current research and theory*, pages 64–99, 1972.

David Ricardo, Francisco Solano Constâncio, and Jean Baptiste Say. *Des principes de l'économie politique et de l'impôt*, volume 2. H. Dumont, 1835.

Erin L Rich and Jonathan D Wallis. Decoding subjective decisions from orbitofrontal cortex. *Nature neuroscience*, 19(7):973–980, 2016.

Hector M Robertson and William L Taylor. Adam smith's approach to the theory of value. *The Economic Journal*, 67(266):181–198, 1957.

Alexandra G Rosati and Brian Hare. Chimpanzees and bonobos distinguish between risk and ambiguity. *Biology letters*, 7(1):15–18, 2011.

Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

Kai Ruggeri, Sonia Alí, Mari Louise Berge, Giulia Bertoldo, Ludvig D Bjørndal, Anna Cortijos-Bernabeu, Clair Davison, Emir Demić, Celia Esteban-Serna, Maja Friedemann, et al. Replicating patterns of prospect theory for decision under risk. *Nature human behaviour*, 4(6):622–633, 2020.

Aldo Rustichini. Neuroeconomics: what have we found, and what should we search for. *Current Opinion in Neurobiology*, 19(6):672–677, December 2009. ISSN 1873-6882. doi: 10.1016/j.conb.2009.09.012.

Yutaka Sakai and Tomoki Fukai. The actor-critic learning is behind the matching law: matching versus optimal behaviors. *Neural computation*, 20(1):227–251, 2008.

P. A. Samuelson. A Note on the Pure Theory of Consumer's Behaviour. *Economica*, 5(17):61–71, 1938. ISSN 0013-0427. doi: 10.2307/2548836. URL <https://www.jstor.org/stable/2548836>.

Emily N Satinsky, Tomoki Kimura, Mathew V Kiang, Rediet Abebe, Scott Cunningham, Hedwig Lee, Xiaofei Lin, Cindy H Liu, Igor Rudan, Srijan Sen, et al. Systematic review and meta-analysis of depression, anxiety, and suicidal ideation among ph. d. students. *Scientific Reports*, 11(1):1–12, 2021.

References

- Leonard J. Savage. *The foundations of statistics*. Courier Corporation, 1954. ISBN 0-486-62349-1.
- Geoffrey Schoenbaum and Matthew Roesch. Orbitofrontal cortex, associative learning, and expectancies. *Neuron*, 47(5):633–636, 2005.
- Geoffrey Schoenbaum, Yuji Takahashi, Tzu-Lan Liu, and Michael A McDannald. Does the orbitofrontal cortex signal value? *Annals of the New York Academy of Sciences*, 1239:87, 2011.
- Tom Schonberg and Leor N Katz. A neural pathway for nonreinforced preference change. *Trends in Cognitive Sciences*, 24(7):504–514, 2020.
- Ernst Schröder. *Lehrbuch der Arithmetik und Algebra für Lehrer und Studierende*, volume 1. BG Teubner, 1873.
- Wolfram Schultz, Peter Dayan, and P. Read Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997. ISBN: 0036-8075 Publisher: American Association for the Advancement of Science.
- Odelia Schwartz, Anne Hsu, and Peter Dayan. Space and time in visual context. *Nature Reviews Neuroscience*, 8(7):522–535, 2007.
- Amartya Sen. Behaviour and the concept of preference. *Economica*, 40(159):241–259, 1973.
- David Teira Serrano. A positivist tradition in early demand theory. *Journal of Economic Methodology*, 13(1):25–47, 2006.
- Guillaume Sescousse, Xavier Caldú, Bàrbara Segura, and Jean-Claude Dreher. Processing of primary and secondary rewards: A quantitative meta-analysis and review of human functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 37(4):681–696, May 2013. ISSN 01497634. doi: 10.1016/j.neubiorev.2013.02.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S0149763413000377>.
- George LS Shackle. A non-additive measure of uncertainty. *The Review of Economic Studies*, 17(1):70–74, 1949.
- Shepard Siegel and Lorraine G Allan. The widespread influence of the rescorla-wagner model. *Psychonomic Bulletin & Review*, 3(3):314–321, 1996.
- Herbert A Simon. *Administrative behavior*. Simon and Schuster, 1947.
- Herbert A. Simon. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99–118, February 1955. ISSN 0033-5533. doi: 10.2307/1884852. URL <https://academic.oup.com/qje/article/69/1/99/1919737>. Publisher: Oxford Academic.
- Herbert A Simon. Theories of bounded rationality. *Decision and organization*, 1(1):161–176, 1972.
- Herbert A Simon. Invariants of human behavior. *Annual review of psychology*, 41(1):1–20, 1990.
- Burrhus Frederic Skinner. *The behavior of organisms: An experimental analysis*. D. Appleton-Century Company, incorporated, 1938.

- Burrhus Frederic Skinner. A case history in scientific method. *American psychologist*, 11(5):221, 1956.
- Paul Slovic. The construction of preference. *American psychologist*, 50(5):364, 1995.
- Paul Slovic and Sarah Lichtenstein. Comparison of bayesian and regression approaches to the study of information processing in judgment. *Organizational behavior and human performance*, 6(6):649–744, 1971.
- Paul Slovic and Sarah Lichtenstein. Preference reversals: A broader perspective. *The American Economic Review*, 73(4):596–605, 1983.
- Adam Smith. *The wealth of nations [1776]*, volume 11937. na, 1776.
- J. E. R. Staddon and Y. Niv. Operant conditioning. *Scholarpedia*, 3(9):2318, 2008. doi: 10.4249/scholarpedia.2318. revision #91609.
- Kyle Stanford. Underdetermination of Scientific Theory. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021.
- Caleb E Strait, Brianna J Sleezer, and Benjamin Y Hayden. Signatures of value comparison in ventral striatum neurons. *PLoS biology*, 13(6):e1002173, 2015.
- Leo P Sugrue, Greg S Corrado, and William T Newsome. Matching behavior and the representation of value in the parietal cortex. *science*, 304(5678):1782–1787, 2004.
- Gaurav Suri, James J Gross, and James L McClelland. Value-based decision making: An interactive activation perspective. *Psychological review*, 127(2):153, 2020.
- Roland E Suri and Wolfram Schultz. Temporal difference model reproduces anticipatory neural activity. *Neural computation*, 13(4):841–862, 2001.
- Richard S Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.
- Richard S Sutton and Andrew G Barto. Toward a modern theory of adaptive networks: expectation and prediction. *Psychological review*, 88(2):135, 1981.
- Richard S Sutton and Andrew G Barto. A temporal-difference model of classical conditioning. In *Proceedings of the ninth annual conference of the cognitive science society*, pages 355–378. Seattle, WA, 1987.
- Richard S Sutton and Andrew G Barto. Time-derivative models of pavlovian reinforcement. 1990.
- Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018. ISBN 0-262-35270-2.
- Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pages 1057–1063, 2000.

References

- Yuji Takahashi, Geoffrey Schoenbaum, and Yael Niv. Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. *Frontiers in neuroscience*, 2:14, 2008.
- Richard H Thaler, Amos Tversky, Daniel Kahneman, and Alan Schwartz. The effect of myopia and loss aversion on risk taking: An experimental test. *The quarterly journal of economics*, 112(2):647–661, 1997.
- Edward L. Thorndike. Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, 2(4):i–109, 1898. ISSN 0096-9753(Print). doi: 10.1037/h0092987. Place: US Publisher: The Macmillan Company.
- Gerhard Tintner. A Contribution to the Non-Static Theory of Choice. *The Quarterly Journal of Economics*, 56(2):274, February 1942. ISSN 00335533. doi: 10.2307/1881933. URL <https://academic.oup.com/qje/article-lookup/doi/10.2307/1881933>.
- Philippe N Tobler, Christopher D Fiorillo, and Wolfram Schultz. Adaptive coding of reward value by dopamine neurons. *Science*, 307(5715):1642–1645, 2005.
- Edward C Tolman. Cognitive maps in rats and men. *Psychological review*, 55(4):189, 1948.
- S. M. Tom, C. R. Fox, C. Trepel, and R. A. Poldrack. The Neural Basis of Loss Aversion in Decision-Making Under Risk. *Science*, 315(5811):515–518, January 2007a. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1134239. URL <https://www.sciencemag.org/lookup/doi/10.1126/science.1134239>.
- Sabrina M Tom, Craig R Fox, Christopher Trepel, and Russell A Poldrack. The neural basis of loss aversion in decision-making under risk. *Science*, 315(5811):515–518, 2007b.
- Léon Tremblay and Wolfram Schultz. Relative reward preference in primate orbitofrontal cortex. *Nature*, 398(6729):704–708, 1999.
- Christopher Trepel, Craig R Fox, and Russell A Poldrack. Prospect theory on the brain? toward a cognitive neuroscience of decision under risk. *Cognitive brain research*, 23(1):34–50, 2005.
- Alexandre Truc. Forty years of behavioral economics. *Available at SSRN 3762621*, 2021.
- Amos Tversky and Daniel Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232, 1973.
- Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.
- Amos Tversky and Daniel Kahneman. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.
- Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323, 1992a.

- Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, October 1992b. ISSN 1573-0476. doi: 10.1007/BF00122574. URL <https://doi.org/10.1007/BF00122574>.
- Amos Tversky and Eldar Shafir. Choice under conflict: The dynamics of deferred decision. *Psychological science*, 3(6):358–361, 1992.
- Amos Tversky, Shmuel Sattath, and Paul Slovic. Contingent weighting in judgment and choice. *Psychological review*, 95(3):371, 1988.
- William R Uttal. *The new phrenology: The limits of localizing cognitive processes in the brain*. The MIT press, 2001.
- Francisco J Varela, Eleanor Rosch, and Evan Thompson. *The embodied mind: Cognitive science and human experience*. MIT press, 1992.
- Hal R Varian. Revealed preference. *Samuelsonian economics and the twenty-first century*, pages 99–115, 2006.
- William Vickrey. Measuring marginal utility by reactions to risk. *Econometrica: Journal of the Econometric Society*, pages 319–333, 1945.
- Jacob Viner. The utility concept in value theory and its critics. *Journal of Political Economy*, 33(6):638–659, 1925.
- Ivo Vlaev, Nick Chater, Neil Stewart, and Gordon D. A. Brown. Does the brain calculate value? *Trends in Cognitive Sciences*, 15(11):546–554, November 2011. ISSN 1879-307X. doi: 10.1016/j.tics.2011.09.008.
- Andreas Voigt. Zahl und mass in der ökonomik. eine kritische untersuchung der mathematischen methode und der mathematischen preistheorie. *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics*, (H. 4):577–609, 1893.
- J. Von Neumann and O. Morgenstern. *Theory of games and economic behavior*. Theory of games and economic behavior. Princeton University Press, Princeton, NJ, US, 1944.
- Peter BM Vranas. Gigerenzer’s normative critique of kahneman and tversky. *Cognition*, 76(3):179–193, 2000.
- Abraham Wald. Asymptotic minimax solutions of sequential point estimation problems. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 1–11. University of California Press, 1951.
- Léon Walras. *Éléments d’économie politique pure, ou, Théorie de la richesse sociale*. F. Rouge, 1896.
- Leon Walras. Economics and mechanics. *Economics as Discourse*. Boston: Kluwer Academic Publishers. First Published in, 1909.
- Christopher John Cornish Hellaby Watkins. Learning from delayed rewards. 1989.

References

- John B Watson and Rosalie Rayner. Conditioned emotional reactions. *Journal of experimental psychology*, 3(1):1, 1920.
- Bernd Weber, Andreas Aholt, Carolin Neuhaus, Peter Trautner, Christian E Elger, and Thorsten Teichert. Neural evidence for reference-dependence in real-market-transactions. *Neuroimage*, 35(1):441–447, 2007.
- Elke U. Weber, Sharoni Shafir, and Ann-Renee Blais. Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. *Psychological review*, 111(2):430, 2004. ISBN: 1939-1471 Publisher: American Psychological Association.
- Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25(A):287–298, 1988.
- Bernard Widrow and Marcian E Hoff. Adaptive switching circuits. Technical report, Stanford Univ Ca Stanford Electronics Labs, 1960.
- Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256, 1992.
- Robert C Wilson and Anne GE Collins. Ten simple rules for the computational modeling of behavioral data. *Elife*, 8:e49547, 2019.
- Robert C Wilson, Yuji K Takahashi, Geoffrey Schoenbaum, and Yael Niv. Orbitofrontal cortex as a cognitive map of task space. *Neuron*, 81(2):267–279, 2014.
- F. Woergoetter and B. Porr. Reinforcement learning. *Scholarpedia*, 3(3):1448, 2008. doi: 10.4249/scholarpedia.1448. revision #127590.
- Stanley Wong. The” f-twist” and the methodology of paul samuelson. *The American economic review*, 63(3):312–325, 1973.
- Chris Woolston. Phds: the tortuous truth, 2019.
- Dirk U. Wulff, Max Mergenthaler-Canseco, and Ralph Hertwig. A meta-analytic review of two modes of learning and the description-experience gap. *Psychological bulletin*, 144(2):140, 2018. ISBN: 1939-1455 Publisher: American Psychological Association.
- Eric R. Xu and Jerald D. Kralik. Risky business: rhesus monkeys exhibit persistent preferences for risky options. *Frontiers in Psychology*, 5, April 2014. ISSN 1664-1078. doi: 10.3389/fpsyg.2014.00258. URL <http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00258/abstract>.
- Juliana Yacubian, Jan Gläscher, Katrin Schroeder, Tobias Sommer, Dieter F Braus, and Christian Büchel. Dissociable systems for gain-and loss-related value predictions and errors of prediction in the human brain. *Journal of Neuroscience*, 26(37):9530–9537, 2006.
- Seng Bum Michael Yoo and Benjamin Yost Hayden. Economic choice as an untangling of options into actions. *Neuron*, 99(3):434–447, 2018.



Appendix

Introduction

During my master's degree, I submitted two theses, which were revised, submitted, and eventually published as articles when I was pursuing my doctoral thesis. Both articles intended to test the predictions of certain economic models when implemented under laboratory conditions, i.e. when the conditions for classical rationality are not met. Interestingly, these papers illustrate well the dichotomy between value-first and value-free models. Depending on the experimental conditions, both were useful tools to account for subjects' behavior.

In a first study, we experimentally tested Iwai's model of money emergence (Iwai, 1997). This model attempts to formalize the conditions under which a commodity money can emerge from interactions located in a barter economy, with 3 or 4 goods in circulation. For each good, there is a type of agent that produces it, and one that consumes it. Agents are thus specialized both in production and consumption. The goal for an agent is to obtain its consumption good, which constitutes the reward. Eventually, if the economy converges, one the good will be used as a medium of exchange (in order to avoid frictions), that is, a commodity money. This framework involves multi-step decisions, in the sense that an agent has to predict that it will be obtaining its consumption good by (1) exchanging its production against a transition good (2) then use this good to obtain its consumption good. In addition, in a 3 goods economy, each agent is presented, at each trial, with 2 options (3 goods minus the one already produced and owned). Conversely, in a 4 goods, there are three options. Thus, it implies to solve both the *prediction problem* and the *control problem* (see chapter 2). In humans, we observed that a simple reinforcement learning model best-fitted subjects' behavior, which made sense with regards to the environmental structure.

In a second study, we experimentally tested another economic model, the Hotelling's model (Hotelling, 1929). This model assumes a linear city, where two firms are competing for a market. Consumers are evenly distributed on the segment, and aim to maximize their utility. In order to maximize their profit, firms can vary two parameters: their price and their spatial location. Because of the curse of dimensionality (Bellman, 1956) that results from the combinatorial explosion of decision variables, such environment is not tractable by value-learning models, or at least classical reinforcement learning. Furthermore, we observed that humans' decisions were best accounted by heuristics, which allowed satisfactory performance while reducing the environmental complexity.

A.1 Coordination over a unique medium of exchange under information scarcity






ARTICLE

<https://doi.org/10.1057/s41599-019-0362-2>

OPEN

Coordination over a unique medium of exchange under information scarcity

Aurélien Nioche ^{1,2,3,4,5,13*}, Basile Garcia^{4,5,6,7,8,13}, Germain Lefebvre^{6,7,9,10}, Thomas Boraud^{4,5,11}, Nicolas P. Rougier ^{4,5,8,12,13} & Sacha Bourgeois-Gironde ^{2,3,9,13}

ABSTRACT Several micro-founded macroeconomic models with rational expectations address the issue of money emergence, by characterizing it as a coordination game. These models have in common the use of agents who dispose of perfect or near-perfect information on the global state of the economy and who display full-fledged computational abilities. Several experimental studies have shown that a simple trial-and-error learning process could constitute an explanation for how agents coordinate on a single mean of exchange. However, these studies provide subjects with full information regarding the state of the economy while restricting the number of goods in circulation to three. In this study, by the mean of multi-agent simulations and human experiments, we test the hypothesis according to which coordination over a unique medium of exchange is possible in the context of information scarcity. In our experimental design, subjects and artificial agents are only aware of the outcome of their own decisions. We provide results for economies with 3 and 4 goods to evaluate to which extent it is possible to generalize results obtained with 3 goods to n goods. Our findings show that in an economy à la Iwai, commodity money can emerge under drastic information restrictions with three goods in circulation, but generalization to four or more goods is not guaranteed.

¹Department of Communications and Networking, School of Electrical Engineering, Aalto University, 02150 Espoo, Finland. ²Département d'Etudes Cognitives, Institut Jean Nicod, ENS, EHESS, PSL Research University, 75005 Paris, France. ³Institut Jean Nicod, CNRS, UMR 8129, Paris, France. ⁴Institut des Maladies Neurodégénératives, Université de Bordeaux, 33000 Bordeaux, France. ⁵Institut des Maladies Neurodégénératives, CNRS, UMR 5293, Paris, France. ⁶Laboratoire de Neurosciences Cognitives Computationnelles, Département d'Etudes Cognitives, ENS, PSL Research University, 75005 Paris, France. ⁷Laboratoire de Neurosciences Cognitives Computationnelles, INSERM, U960, Paris, France. ⁸Inria Bordeaux Sud-Ouest, 33405 Talence, France. ⁹Laboratoire d'Economie Mathématique et de Microéconomie Appliquée, Université Panthéon Assas, 75006 Paris, France. ¹⁰Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. ¹¹Centre Expert Parkinson, CHU Bordeaux, 33000 Bordeaux, France. ¹²LaBRI, Université de Bordeaux, INP, CNRS, UMR 5800, 33405 Talence, France. ¹³These authors contributed equally: Aurélien Nioche, Basile Garcia, Nicolas P. Rougier, Sacha Bourgeois-Gironde
*email: nioche.aurelien@gmail.com

Introduction

In the last decades, monetary economics has shifted from a purely macroeconomic understanding of money to an analysis of its micro-foundations, both in its game-theoretical and behavioral dimensions. Following the intuitions of Karl Menger (1892) and starting with the Jones' model in the mid-1970's (Jones, 1976), several *search-theoretic* models have been proposed in order to identify the conditions for money emergence (Diamond, 1984; Kiyotaki and Wright, 1989, 1991; Oh, 1989; Aiyagari and Wallace, 1991; Kiyotaki and Wright, 1993; Shi, 1995; Iwai, 1996; Kehoe et al., 1993; Wright, 1995; Luo, 1998). They are considered search-theoretic models in the sense that they describe situations where agents need to search for a trading partner before transacting (Nosal and Rocheteau, 2011). Besides, these models belong to the class of micro-founded macroeconomic models with rational expectations. Agents with rational expectations can take advantage of all the available information to form their expectations and decide which action is optimal on the belief that every other agent in the economy has a similar ability (Muth, 1961).

Their first advantage is that they explain a macroeconomic phenomenon—money emergence—from individual decision-making processes. The second advantage of these models is that they explain money emergence that does not require the economies to be centralized: they do not need to assume a monetary authority for the agents to coordinate over a unique medium of exchange. Focusing on the function of a medium of exchange, these models highlight the key role that the money can play in limiting frictions in exchange processes (i.e., the difficulty to find an exchange partner). However, these models are based on three unrealistic assumptions: the omniscience of economic agents, infinite time and an extremely large number of agents (unbounded).

A question that immediately arises is whether money emergence without a monetary authority is possible in an economy populated by agents with restricted abilities and having limited access to information. More precisely, we want to know whether coordination over a unique medium of exchange is possible when agents proceed by trial and error and have access to local information only.

A partial answer has been brought to this question, through agent-based simulations with artificial agents using a reinforcement learning process (Marimon et al., 1990; Duffy and Ochs, 1999; Kindler et al., 2017) in a Kiyotaki-and-Wright's environment (Kiyotaki and Wright, 1989, 1993). In these simulations, reinforcement learning agents have by construction limited computational abilities, and their informational inputs are only constituted by the success and failures of each exchange attempt. In contrast to Kiyotaki-and-Wright's theoretical agents, they are completely blind to the global state of the economy, and the tuning of their preferences does not rely on the knowledge of the latter. Yet, results report achievement of monetary equilibria, indicating that fully rational agents are not required for money to emerge. In a similar perspective, other work considers the question of money emergence under heterogeneous beliefs, where some agents are rational, and the remaining fraction learns by an adaptive learning rule, showing that coordination is also eventually possible in this setting (Branch and McGough, 2016).

The Kiyotaki's and Wright's model (Kiyotaki and Wright, 1989, 1993) has been experimentally tested, to show if results obtained analytically or by numerical simulation were reproducible with actual human subjects. It had been shown that a monetary equilibrium can be reached with human subjects evolving in a search-theoretic environment (Brown, 1996; Duffy and Ochs, 1999; Duffy, 2001), or at least reaching a high proportion of speculators (Lefebvre et al., 2018). Interestingly, it has

been shown that a reinforcement model fits well their experimental data obtained in a Kiyotaki-and-Wright's environment (Kiyotaki and Wright, 1989, 1993), suggesting that although more sophisticated behavior rules were available, subjects tended to favor immediate past feedback (Duffy and Ochs, 1999; Duffy, 2001).

One first critic that we can address the computational and experimental aforementioned studies, is that although they succeeded in demonstrating achievements of monetary equilibrium, they were mainly considering the fundamental equilibrium of Kiyotaki and Wright (Kiyotaki and Wright, 1989, 1993). Indeed, Kiyotaki and Wright (Kiyotaki and Wright, 1989, 1993) consider two types of equilibrium: (i) *fundamental*, where the monetized good is less costly to store than the other goods in circulation, what explains easily why it is preferred, (ii) *speculative*, where some agents are required to incur at first supplementary costs (i.e., to speculate). The *speculative* equilibrium is particularly interesting, as it provides insight about a specific cognitive ability that could sustain money emergence (i.e., the ability to endorse a cost on short term with view on distant goals), and yet, it is the one for which the results are the scarcest (Brown, 1996; Duffy and Ochs, 1999; Duffy, 2001; Kindler et al., 2017; Lefebvre et al., 2018). Secondly, in contrast with virtual agents learning by reinforcement that are only provided with scarce information, human subjects had access to information about the global state of the economy in studies mixing the use of artificial agents and human subjects (Duffy and Ochs, 1999; Duffy, 2001; Lefebvre et al., 2018). Thirdly, to our knowledge, these computational and experimental studies are based on search-theoretic models involving only three goods (Brown, 1996; Duffy and Ochs, 1999; Duffy, 2001; Kindler et al., 2017; Lefebvre et al., 2018). In this case, only one type of agent uses the monetary good genuinely as a medium of exchange. It remains to know whether their conclusions can hold if there are more than three goods in circulation.

Let us note that in recent literature, numerous questions have been treated through an experimental money-emergence paradigm: Whether a convergence on a money equilibria is preferred to a gift exchange equilibria, where an agent has the possibility to give a good in the hope of obtaining another later (Duffy and Puzzello, 2014), how inflation tax affects economic activity (Anbarci et al., 2015), how a foreign money may be accepted by agents in an international framework (Jiang and Zhang, 2018; Ding et al., 2018), how a monetary equilibrium is reachable under assumption of a finite horizon (Davis et al., 2019), or even how when a first money already circulates in the economy, a second may emerge (Rietz, 2019). However, either they assume a central authority that injects money (Anbarci et al., 2015; Ding et al., 2018), either money does not emerge endogenously, as a fraction of agents is first provided with tokens (worthless goods that none agents consume) they are compelled to exchange to obtain their consumption good (Duffy and Puzzello, 2014; Jiang and Zhang, 2018; Davis et al., 2019; Rietz, 2019). In these experiments, the cognitive requirements for money emergence as an endogenous process are thus never explicitly tested.

The purpose of this study is to know whether economies populated with human subjects can reach a monetary state in the context of information scarcity, that is in a case of extremely *incomplete* information in the sense of the game theory, forcing the subjects to take their decisions under a strong form of *ambiguity*. More precisely, this study aims to investigate whether coordination over a unique medium of exchange can occur with subjects only experiencing the direct outcome of their decision, learning by trial-and-error and without any additional information.

Hence, the question is to know whether results obtained with virtual agents combining a restriction on computational abilities and informational input can be generalized to economies populated with humans. To assess their reliability and to broaden our conclusions, we decided to include an additional good, including in our study economies with four goods in circulation. To meet these goals, we borrowed certain elements from the previous search-theoretical models to define the structure of our economies, such as the production-consumption specialization and the absence of double coincidence of wants (i.e., if an agent produces i and consumes j , no agent produces j and consumes i , so that pure bartering is not an effective solution). However, instead of using an environment a la Kiyotaki and Wright (Kiyotaki and Wright, 1989, 1993), we decided to use a search-theoretical structure that presents more generality than Kiyotaki and Wright’s one, based on the Iwai’s model (Iwai, 1996). Iwai’s model differs in two fundamental ways from Kiyotaki and Wright’s model (Kiyotaki and Wright, 1989, 1993): (i) the *exchange technology* consists in random pairing inside markets specialized in a pair of good while in Kiyotaki and Wright (1989, 1993), agents are randomly matched regardless of any other characteristic (ii) there are no *storage cost*, such as storing a good i is not costlier than storing good j . That is why we decided to adopt an Iwai-like environment with indistinguishable goods, in a way to avoid that money emergence bears on intrinsic features of goods, as it is in the case of the Kiyotaki and Wright’s *fundamental* equilibrium. We began by conducting a series of simulations. In the simulated economies, agents are producing a certain good and looking to obtain another one through exchanges, have little knowledge about the environment in which they operate—they only know if their attempt of exchange was a success or a failure. They are learning using a basic reinforcement mechanism, associating a value to each choice option available to them and updating by trial-and-error the efficiency of each type of exchange. We used the results of these simulations to identify the experimental conditions that would promote the coordination over a single medium of exchange. Subsequently, we observed the behaviors of human subjects under similar informational constraints and we compare the theoretical and experimental results. To conclude, we discuss the possibility of coordination over a unique medium exchange in the context of information scarcity, in a three and four goods setting.

Materials and methods

Model

General framework. Each economy is composed of different types of agents. A type of agent is defined by what agents of this type produce and consume. The goal of each agent is to obtain his consumption good. Agents proceed to exchanges between them to achieve this goal. Agents have feedback only about their exchange attempt and learn by reinforcement the efficiency of each type of exchange. We vary across simulations the distribution of agents among the existing types. By construction, if a good m becomes money, an agent that produces it or consumes it should try to exchange directly his production good against his consumption good. Otherwise, the agent is supposed to use it as a medium of exchange, that is to exchange his production good against m , and then m against his consumption good.

Production-consumption specialization. We consider an economy with G goods in circulation, with $G \geq 3$. We denote these goods $1, 2, \dots, G$. Each agent is specialized in production and consumption. A agent of type (i, j) produces good i and consumes good j (with $j \neq i$). We suppose a non double coincidence of needs: if an agent of type (i, j) exists, then an agent of type (j, i)

does not exist. We use a *minimally connected endowment-need distribution* (Iwai, 1996), such that existing agent types are: $(G, 1), (1, 2), \dots, (G - 1, G)$. The number of agents for each type is exogenously set. We designate by x_{G1} the number of agent of type $(G, 1)$, x_{12} the number of agent of type $(1, 2)$, ..., x_{G-1G} the number of agent of type $(G - 1, G)$. Each agent enters the economy equipped with a unit of its production good. Each time an agent receives its consumption good, it consumes it and immediately after, produces a new unit of its production good (each agent owns a single storage unit).

Exchange technology. The exchange technology relies on a *trading-post mechanism* (Iwai, 1996). At each time step, each agent chooses the type of exchange it wants to perform, depending on the good it has in hand. This choice determines to which market it goes. There is an equal number of markets and goods in circulation. Each market is specialized in a pair of good (i, j) , such as in the ij -market it is possible to exchange i against j , and j against i . Our trading technology works synchronously (i.e., all exchanges occur simultaneously). Thus, in each ij -market, we randomly associates each i -seller – j -buyer to a j -seller – i -buyer, if there is a sufficient number of j -sellers– i -buyers. Therefore, in each ij -market, the probability of successfully exchanging a good i against a good j depends on the respective number of i -sellers – j -buyers and j -sellers– i -buyers (e.g., if there is in the ij -market at time t , 4 i -sellers – j -buyers and 8 j -sellers – i buyers, 4 ij -exchanges will take place and the probability of success for a i -seller – j buyers is 0.5 while a j -seller – i -buyer will proceed to the desired exchange with certainty).

Information scarcity. An agent does not know other agents’ choices, nor the probabilities of success of each exchange: the only information it has access to is whether or not it succeeded in the desired exchange.

Strategies. The goal of each agent is to obtain as quickly as possible his consumption good.

We will specifically consider:

- The direct exchange strategy. For a type- ij agent with i in hand (his production good), it consists of trying an exchange against j (his consumption good).
- The indirect exchange strategy with k as a medium of exchange. For a type- ij agent with i in hand (his production good), it consists of trying an exchange against the good k (with $k \neq i, j$). With k in hand, it consists of trying an exchange against j (his consumption good).

Simulations

Decision-making process. Each agent learns to estimate the success rate of each type of exchange. This allows it to estimate the time needed to get its consumption good depending on the choice is made.

Success rate estimates for each exchange type are based on a reinforcement learning process. At time step t , when an agent attempts to exchange i against j , it updates the success rate estimation associated to the exchange of type (i, j) , noted e_{ij} according to:

$$e_{ij}^{t+1} = e_{ij}^t + \alpha \cdot (s - e_{ij}^t) \tag{1}$$

with $\alpha \in [0, 1]$, a free parameter and s , a binary variable such as $s = 1$ if the agent succeeded in his exchange, 0 otherwise. α is a learning rate which defines to which extent an agent takes into account his latest attempted exchange. If $\alpha = 1$, the agent considers only his latest attempted exchange. If $\alpha = 0$, the agent

does not take into account the new observations of failure or success of the last attempted exchange.

When making a decision, each agent considers the expected temporal interval between the time of choice and the time he gets his consumption good. It is assumed that the longer the time interval, the lower the value for the agent. Let $v(ij)$ be the value associated to the choice ij (i.e., exchange i against j) and Δ_{ij} the estimation by the agent of the time that will be spent before consumption if he chooses ij :

$$v(ij) = 1/(1 + \beta)^{\Delta_{ij}} \quad (2)$$

with $\beta > 0$, a free parameter. β is a discount factor parameter: the closer to 0, the more subjective values are discounted with time (Osborne, 2016). Since it takes at least one unit of time for the agent to get its consumption good, the value function v is bounded between 0 and 1.

We assume that for each exchange of type (i, j) , the agent has an estimation of the success rate associated to this type of exchange (e_{ij}). The higher the estimated success rate, the lower the estimated time to succeed in this exchange. Let δ_{ij} be the estimated time to achieve a type- ij exchange:

$$\delta_{ij} = 1/e_{ij} \quad (3)$$

For a type- ij agent, $\Delta_{ij} = \delta_{ij}$. If a type- ik agent (with $k \neq j$), the value of Δ_{ij} depends on the action policy chosen by the agent, as Δ_{ij} would be equal in this case to the sum of the δ -values for each intermediary exchange planned by the agent. For instance, for a type- ik agent following an indirect exchange strategy with good j , $\Delta_{ij} = \delta_{ij} + \delta_{jk}$. An exhaustive description of valuation functions for the specific case where $G = 3$ is given in the supplementary section.

Agents make decisions using a probabilistic decision rule. The standard approach is to use a softmax function to introduce stochasticity in choice (Sutton and Barto, 1998). However, Apesteguia and Ballester (2018) show that the combination of a softmax rule and either a risk-sensitivity or a temporal discounting model can be problematic, as the parameter describing the risk-sensitivity discounting effect can have a non-monotonic effect on the variable of interest. For this reason, the rule implemented is a simple ϵ -rule (Sutton and Barto, 1998). Let $v(ij)$ be the value associated with choice ij and $p(ij)$, the probability to choose to exchange i against j . $p(ij)$ is computed as follows:

$$p(ij) = \begin{cases} 1 - \gamma & \text{if } \forall k : v(ij) > v(ik), \\ \gamma/(G - 1) & \text{otherwise.} \end{cases} \quad (4)$$

with $\gamma \in [0, 1]$, a free parameter. γ is an exploitation-exploration rate (Sutton and Barto, 1998): the lower the γ -value, the more prone the agent will be to choose the option with the highest subjective value. On the contrary, the higher the γ -value, the more the agent will be prone to choose another option.

Protocol and parametrization. We ran 10,800 simulations with $G = 3$ and 10,800 simulations with $G = 4$. Each simulation lasted 100 time-steps. The exploration parameter (ϵ) was varied between 0.10 and 0.15. The learning rate (α) was varied between 0.10 and 0.25. The discount factor (β) was varied between 0.80 and 1.20. The initial values of success rate estimates for all types of exchange were set to 1. The fact that the initial values were set to 1 precluded the presence of bias in preferences (such as bias such as the appearance of commodity money was more likely). With these values, the value associated with exchanging his production good against his consumption good was indeed higher than the value of any other exchange for all agents, implying that all agents were preferring the direct exchange strategy at the first time-step.

When $G = 3$, x_{31} was set to 50 while x_{12} and x_{23} were varied between 10 and 200.

When $G = 4$, x_{41} and x_{12} were set to 50 (following results from simulations with $G = 3$) while x_{23} and x_{34} were varied between 10 and 200.

Artificial experiments. We ran 4 separate simulations before the experiment using the same distribution of agents as for experiments (2 matching the conditions of Experiment I and 2 matching the conditions of Experiment II). The cognitive parametrization of the agents was: $\alpha = 0.175$, $\beta = 1.000$ & $\gamma = 0.125$ (these values correspond to the average value of each parameter used for the simulations).

Post-hoc simulations. We fitted our behavioral data on the decision-making model using Scipy's (Jones et al., 2001) differential evolution algorithm (provided by the module *optimize*). We optimized model parameters by minimizing the negative log-likelihood of the model for each subject individually.

Using the best-fit parameter values of the subjects to parametrize the artificial agents (the distribution of the best-fit parameter values is given in Fig. S18A of the Supplementary Section), we ran 4 post-hoc simulations (2 matching the conditions of Experiment I and 2 matching the conditions of Experiment II).

Experiment I

Subjects. Sixty-six subjects have been recruited by the Maison des Sciences Économiques (106–112, boulevard de l'Hôpital, 75013 Paris, France). The ethics approval for this project was provided by the Institutional Review Board of the Paris School of Economics. In line with ethical guidelines, all participants provided their informed consent before proceeding to the experiment and filled in a survey asking their age and gender. Financial compensation of 10 euros was offered to each participant, with a bonus proportional to their score (a subject earned a point when he succeeded to obtain its consumption good and each point corresponded to 0.20 euros). The average reward was 15.41 euros (± 1.80 STD). We noticed a gender parity (women represented 48.5% and men 51.5%). The average age was 29.42 (± 12.55 STD).

Task. A subject plays the role of a producer of a good i and a consumer of a good j , in an economy comprising either 30 (uniform condition) or 36 (non-uniform condition) subjects. During 50 time steps, he has to choose which type of exchange he wants to try, among two options (e.g., with good 1 in hand, he has to choose between trying to exchange good 1 against good 2, or good 1 against good 3). The only information he gets is whether he succeeded or not in the exchange. Further details are provided in the supplementary section.

User interface. Following the assumption that a visually appealing serious-game would increase the subject's engagement (Wanner, 2014; Comello et al., 2016) and induce naturalistic decision-making (Harrison and List, 2004), we chose to design a game-inspired interface instead of a textual interface (see Fig. 1).

Experimental conditions. All goods being identical, we arbitrarily chose the good 1 as the 'target', that is to say, the good that we wanted to see emerge. Following the simulation results, we contrasted two modes of distributions, either promoting the money emergence or precluding it. Each subject went through only one of the two conditions. The conditions differ by the distribution of agents among types.

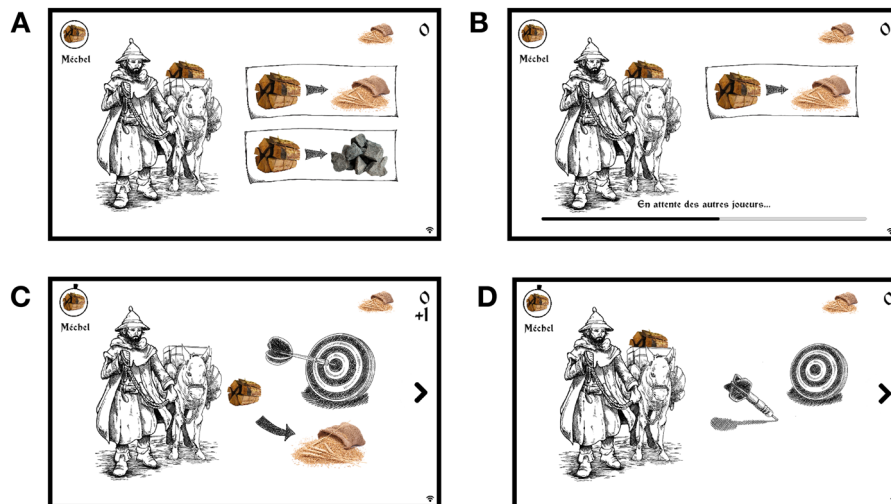


Fig. 1 User interface. Screen-shots corresponding to a 3 good (wood, wheat, and stone) economy. The subject plays the role of a producer of wood, consumer of wheat. **a** Decision-making phase. **b** Waiting screen while other players also take a decision. **c** Successful exchange. **d** Unsuccessful exchange.

- Uniform (U). There is an equal number of agents of each type.
- Non-uniform and promoting the use of a medium of exchange (NUPM). The number of agents for a specific type depends on whether this type involves producing or consuming a specific good, that we arbitrarily chose to be the good 1. The number of agents for a type meeting this condition is half the number of agents of a type not meeting this condition.

The two conditions were the following:

1. $G = 3$ and U-distribution. x_{31} , x_{12} and x_{23} were set to 10.
2. $G = 3$ and NUPM-distribution. x_{31} and x_{12} were equal to 9 but the value of x_{23} was doubled (18)—the choice of setting x_{31} and x_{12} to 9 instead of 10 and x_{23} to 18 instead of 20 is due to the absence of some subjects the day the experiment took place.

Analysis. With three goods in circulation, one type of agent can use the good 1 as a medium of exchange: Agents that produce good 2 and consume good 3. We thus measured for each agent belonging to the type (2, 3), the indirect exchange rate involving good 1. That is the frequency rate at which a subject of type (2, 3) asks for the good 1 to use it as a medium of exchange to get his consumption good 3 from his production good 2. For statistical analysis of the human experiment as well the experiment-like simulations, we averaged this measure overtime for the last third of the trials, to assert learning curves were stable. We then compared these results across uniform and non-uniform distributions of agent types. As we did not expect a normal distribution of data due to clustering effects at the boundaries of our scale, assessment of statistic relevance of our observations has been made with Mann–Whitney’s *U* ranking test (Mann and Whitney, 1947), applying Bonferroni’s corrections for multiple comparisons. We set the significance threshold at 5%.

Experiment II

Subjects. 100 subjects have been recruited under the same conditions as for Experiment I. The remuneration was computed the same way and the average reward was 14.29 euros (± 1.53 STD). We also noticed a gender parity (women represented 50.0% and men 50.0%). The average age was 28.97 years old (± 13.01 STD).

Task. The task is similar to Experiment I, except that they were 4 goods in circulation and that economies were comprising either 40 (uniform condition) or 60 (non-uniform condition) subjects. Also, as a consequence of having 4 goods in circulation, subjects were having 3 alternatives each time, instead of 2 (for instance, with the good 1 in hand, they had a choice between trying to exchange it against the good 2, 3 or 4).

Experimental conditions. As in experiment I, the parametrization of the economies for each condition has been based on the simulation results (see Fig. 2). Hence, the distribution was either uniform (U), either non-uniform promoting the use of a medium of exchange (NUPM):

1. $G = 4$ and U-distribution. x_{41} , x_{12} , x_{23} , x_{34} were set to 10.
2. $G = 4$ and NUPM-distribution. x_{41} , and x_{12} were still equal to 10 but the values of x_{23} and x_{34} were doubled (20).

Analysis. With four goods in circulation, two agent types can use the good 1 as a medium of exchange: Agents that produce good 2 and consume good 3 and agents that produce good 3 and consume good 4. We measured for each agent belonging to the type (2, 3) and (3, 4) the frequency rate at which a subject asks to trade its production good for the good 1 to obtain its consumption good. For statistical analysis of the human experiment as well the experiment-like simulations, we averaged this measure overtime for the last third of the trials, to assert learning curves were stable. We then compared these results across the uniform and non-uniform distribution of agent types. As we did not expect a normal distribution of data due to clustering effects at the boundaries of our scale, assessment of statistic relevance of our observations has been made with Mann–Whitney’s *U* ranking test (Mann and Whitney, 1947), applying Bonferroni’s corrections for multiple comparisons. We set the significance threshold at 5%.

The Supplementary section provides further details, and in particular a summary of the experiment parametrization in Tables S1 and S2.

Results

Simulations

3 goods setting. When $G = 3$, the highest frequency of indirect exchanges with good 1 is observed when the value of x_{31} is equal to that of x_{12} and when the value of x_{23} is at least twice that of x_{31}

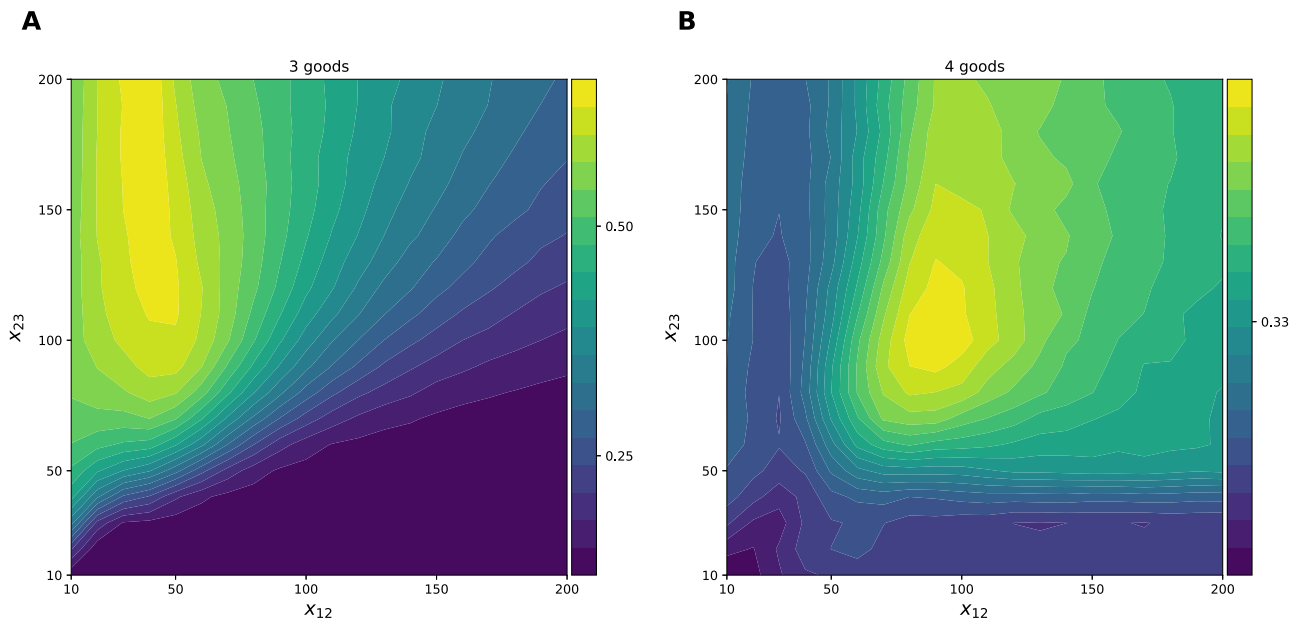


Fig. 2 Simulation: Influence of agents distribution on the use of a medium of exchange. Based on these simulation results, we deduced the optimal experimental conditions required to see money emerge with human subjects. **a** The phase diagram summarizes the results of 10,400 simulations with 3 goods. The number of type (3, 1) agents is set to 50 while the number of agents of type (1, 2) and (2, 3) varies between 10 and 200 (corresponding, respectively, to the values on the x-axis and y-axis). The hotter the color, the higher the indirect exchange frequency involving good 1 as a medium of exchange. In a three goods economy, the highest frequency of indirect exchanges with good 1 observed is when the value x_{12} as well the value of x_{23} is nearly twice that of x_{31} . **b** Similarly, the phase diagram in B panel summarizes the results of 10,400 simulations with 4 goods. The number of agent of types (4, 1) and (1, 2) is set at 50 while the the number of agents of type (2, 3) and (3, 4) varies between 10 and 200 (corresponding, respectively, to the values on the x-axis and y-axis). In a four goods economy, the highest frequency of indirect exchanges with good 1 observed is when the value x_{23} as well the value of x_{34} is nearly twice that of x_{12} and x_{41} .

(see Fig. 2). One may notice that the use of a uniform distribution of agent types ($x_{31} = 50$, $x_{12} = 50$, $x_{23} = 50$) results in a low frequency of indirect exchanges with good 1.

4 goods setting. When $G = 4$, the highest frequency of indirect exchanges with good 1 is observed when the values of x_{23} and x_{34} are nearly twice that of x_{41} and x_{12} (see Fig. 2). The use of a uniform agent type distribution ($x_{41} = 50$, $x_{12} = 50$, $x_{23} = 50$, $x_{34} = 50$) results in a low frequency of indirect exchanges with good 1.

Experimental setup. Put together, these results led us to formulate the following operational hypotheses regarding our experiments: (i) setting the number of one particular type of agents to half of the other agent types promotes the use of its production good as a medium of exchange (ii) setting the number of agents of each type equal precludes the emergence of a medium of exchange.

Hence, for the Experiment I, we set the value of x_{12} equal to that of x_{31} and set the value of x_{23} twice that of x_{31} for the simulations under experimental conditions with $G = 3$ where our goal was to promote money emergence (see Fig. 3). For the Experiment II, we set the value of x_{12} equal to that of x_{41} and to set the value of x_{23} and x_{34} twice that of x_{41} for the simulations under experimental conditions with $G = 4$ where our goal was to promote money emergence (see Fig. 4).

Experiment I

Artificial experiment. To make predictions about the experiment with human subjects, we ran 2 additional simulations, using a parametrization identical to the two experimental conditions (see Table S1). In one of the two conditions, we used a uniform distribution types while in the other, we promoted the use of good 1

as a medium of exchange, by using a non-uniform distribution of agent types (one can note that as all the goods are identical, the choice to promote good 1 is arbitrary).

With $G = 3$ (see Fig. 3), we observe that the median frequency of indirect exchanges with good 1 by agents of type (2, 3) is (i) above chance level, and (ii) significantly greater in the NUPM-distribution than in the U-distribution ($U = 21.0$, $p < 0.001^*$, $n = 28$). This means that agents that neither produce the good 1 nor consume it try to obtain it when they have their production good in the hand and, once in the hand, try to obtain their consumption good using it as an intermediary good.

Human experiment. In line with the results of the simulation, we observe that the median frequency of indirect exchanges with good 1 by subjects of type (2, 3) is (i) above chance level, and (ii) significantly greater in the NUPM-distribution than in the U-distribution ($U = 50.5$, $p = 0.031^*$, $n = 28$).

Post-hoc simulations. The simulations using the best-fit parameter values led to results that have the same pattern as the experimental results. With three goods we observe that the median frequency of indirect exchanges with good 1 by agents of type (2, 3) is significantly greater in the NUPM-distribution than in the U-distribution ($U = 48.0$, $p = 0.023^*$, $n = 28$).

Experiment II

Artificial experiment. To make predictions about the experiment with human subjects, we ran two additional simulations, using a parametrization identical to the two experimental conditions (see Table S2). In one of the two conditions, we used a uniform distribution types while in the other, we promoted the use of good 1 as a medium of exchange, by using a non-uniform distribution of

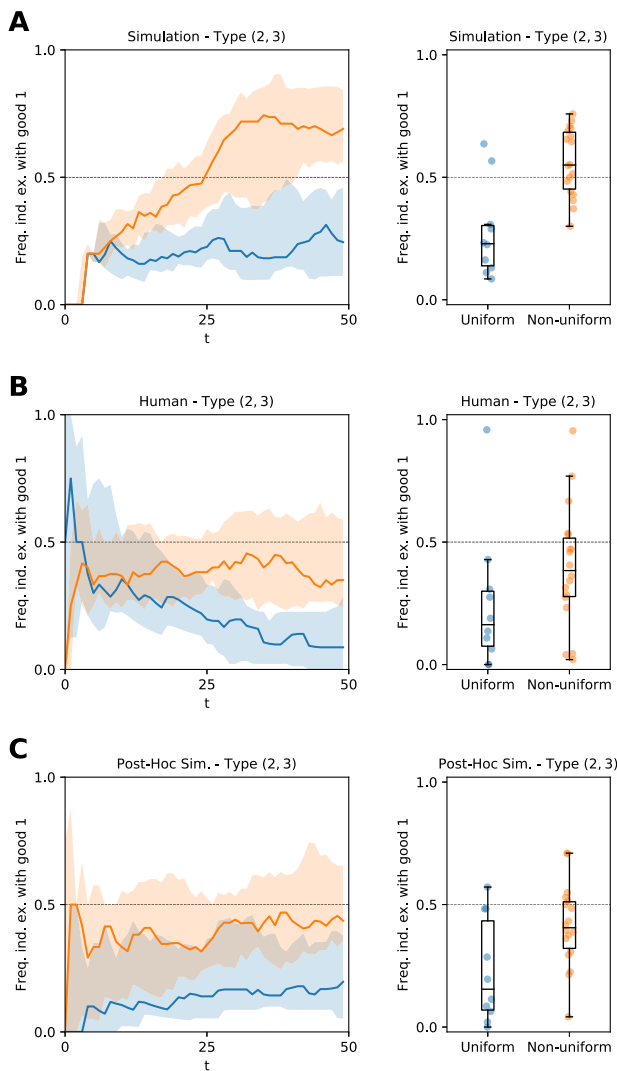


Fig. 3 Experiment I: The use of a medium of exchange with three goods in circulation. We contrast the U-distribution of agent types (blue color) with the NUPM-distribution (orange color). In a three goods economy, only the (2, 3) type of agent can use Good 1 as money. The left side plots represent the moving median (\pm STD) of the frequency of use of a medium of exchange for each individual over time with a 25 time-step window. On the box plots (right side), each dot represents the averaged frequency over time for either one artificial agent (panel **a** and panel **c**), or one human subject (panel **b**) belonging to the (2, 3) agent type. The gray dotted lines indicate the chance level. **a** We observe that in the NUPM-distribution, the median frequency of indirect exchanges involving good 1 is significantly greater than in the U-distribution ($p < 0.05$), showing that the good 1 is used as a medium of exchange significantly more frequently in the NUPM-distribution than in the U-distribution. **b** We replicate this result with human subjects: in the NUPM-distribution, the median frequency of indirect exchanges involving good 1 is significantly greater than in the U-distribution ($p < 0.05$). **c** Running post-hoc simulations with the best-fit parameters of the human subjects, we obtain the same pattern as the experimental results: the median frequency of indirect exchanges involving good 1 is significantly greater than in the U-distribution ($p < 0.05$).

agent types (one can note that as all the goods are identical, the choice to promote good 1 is arbitrary).

With $G = 4$, two types of agent are able to use good 1 as a medium of exchange: (2, 3) and (3, 4). We observe that the median frequency of indirect exchanges with good 1 by (2, 3)

agents (see Fig. 4a) is (i) above chance level, and (ii) significantly greater in the NUPM-distribution than in the U-distribution ($U = 21.0, p < 0.001^*, n = 30$). Similarly, the median frequency of indirect exchanges with good 1 by (3, 4) agents (see Fig. 4b) (i) is above chance level, and (ii) significantly greater in the NUPM-distribution than in the U-distribution ($U = 28.0, p = 0.002^*, n = 30$).

Human experiment. For the condition with $G = 4$, we expected the use of the good 1 as money to be promoted by both agent types (2, 3) and (3, 4). But contrary to what has been observed in the artificial agents, the median frequency of indirect exchanges with good 1 by agents of type (2, 3) (see Fig. 4a) is not significantly greater in the NUPM-distribution than in the U-distribution ($U = 56.0, p = 0.056, n = 30$). Similarly, the median frequency of indirect exchanges with good 1 by agents of type (3, 4) (see Fig. 4b) is not significantly greater in the NUPM-distribution than in the U-distribution ($U = 77.5, p = 0.333, n = 30$).

Post-hoc simulations. The simulations using the best-fit parameters value led to results that have the same pattern as the experimental results. The median frequency of indirect exchanges with good 1 by agents of type (2, 3) is not significantly different in the NUPM-distribution than in the U-distribution ($U = 99.0, p = 0.982, n = 30$), as well as for agents of type (3, 4) ($U = 78.5, p = 0.355, n = 30$).

Supplementary section provides more details for both experiments, in particular a summary of the statistical tests (see Table S3), a short demographic analysis (see Figs S1, S2, and Table S4), the representation of individual behavior (see Figs S3 and S4), a sensitivity analysis to free parameters (see Fig. S5 and Table S5), some post hoc simulations varying some environment parameters and also using alternative decision-making models (see Figs S7, S8, S10–S17, and Tables S7 and S8), more details about the model fitting and a model comparison (see Figs S6, S18, S19, and Tables S6, S9, S10).

Discussion

The results obtained by simulation are in line with our initial assumption: the emergence of commodity money is possible in a decentralized economy with agents endowed with limited computational abilities and having very poor information on the global state of the economy. Indeed, they show that manipulating the agent type distribution is sufficient to foster the emergence of a unique medium of exchange in a 3 goods economy, as well as in a 4 goods economy.

To assess the robustness of these computational results, we conducted two experiments. In contrast to previous experimental studies (Marimon et al., 1990; Duffy, 2001; Kindler et al., 2017), human subjects did not have access to any statistic regarding the current state of the economy in which they were evolving, and in particular the choices of the other participants. The only feedback that they got at each iteration of the game was whether the exchange was successful. Also, contrary to recent experimental studies (Duffy and Puzzello, 2014; Anbarci et al., 2015; Ding et al., 2018; Jiang and Zhang, 2018; Davis et al., 2019; Rietz, 2019), there is no monetary authority, and money emerges endogenously since no good is intrinsically devised to become a medium of exchange.

In the 3 goods setting experiment, the experimental results were consistent with the computational results, the manipulation of the agent type distribution being effective in promoting the use of a unique medium of exchange. Although, in the 4 goods setting experiment, this manipulation turned out to be ineffective. The

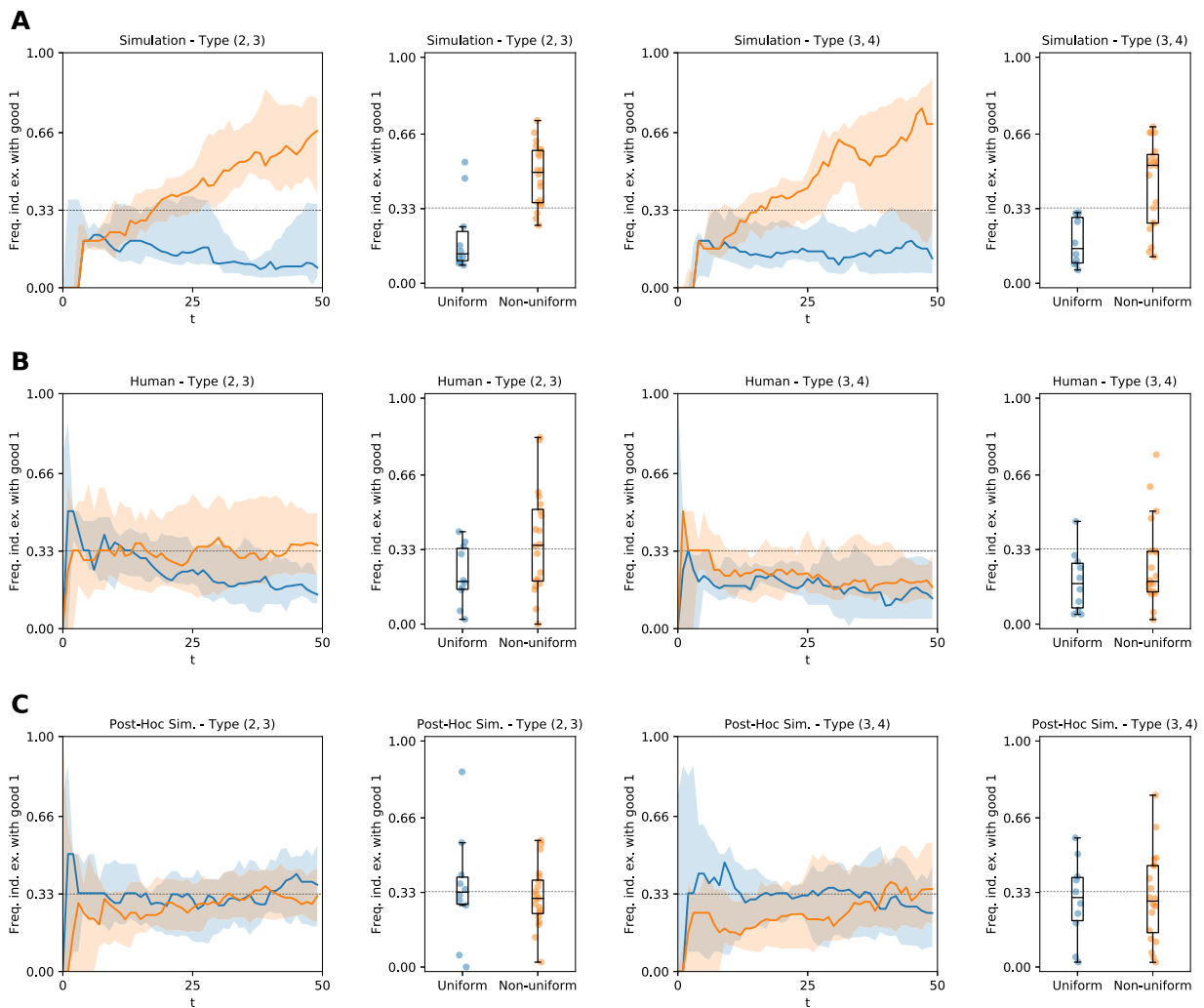


Fig. 4 Experiment II: The use of a medium of exchange with three goods in circulation. We contrast the U-distribution of agent types (blue color) with the NUPM-distribution (orange color). In a four goods economy, two types of agents that can use good 1 as money: (2, 3) and (3, 4). The left side of each pair of plots represents the moving median (\pm STD) of the frequency of use of a medium of exchange for each individual over time with a 25 time-step window. On the box plots (right side), each dot represents the averaged frequency over time for either one artificial agent (panel **a** and panel **c**), or one human subject (panel **b**). Results for (2, 3) agents are depicted on the two leftmost figures of each panel, while results for (3, 4) are depicted on the two rightmost plots. The gray dotted lines indicate the chance level. **a** In simulations and with regards to (2, 3) agents, we observe that in the NUPM-distribution, the median frequency of indirect exchanges involving good 1 is significantly greater than in the U-distribution ($p < 0.05$), showing that good 1 is used a medium of exchange significantly more in the NUPM-distribution than in the U-distribution. Similarly, with artificial agents that belong to the (3, 4) type, we observe that in the NUPM-distribution, the median frequency of indirect exchanges involving good 1 is significantly greater than in the U-distribution ($p < 0.05$). **b** We do not replicate the simulation results from panel **a** with human subjects: the frequency of indirect exchanges with good is not significantly different from the U-distribution ($p > 0.05$). Similarly, we do not replicate the simulation results from panel **b** with human subjects ($p > 0.05$). **c** Running post-hoc simulations with the best-fit parameters of the human subjects, we obtain the same pattern as the experimental results: the median frequency of indirect exchanges involving good 1 are not significantly greater than in the U-distribution for both agent types that are concerned ($p > 0.05$).

results with a 3 goods economy extend precedent works in artificial agents and human using the Kiyotaki and Wright's framework (Marimon et al., 1990; Duffy, 2001; Kindler et al., 2017). In particular, it shows that coordination over a unique medium of exchange is also possible in an Iwai-like environment (Iwai, 1996). Furthermore, it shows that the monetary coordination does not even require agents to have extended knowledge of other players' preferences or to construct a sophisticated belief system: a trial and error approach—in our case, a simple reinforcement learning mechanism—is sufficient. Of course, this coordination between agents over a unique medium of exchange is not systematic: our results suggest that structural constraints are necessary, such as a non-equal distribution of agents over types in

our environment. This can be interpreted as the fact that a particular endowment-need distribution can render sensitive the benefits of coordinating on a unique medium of exchange, thus highlighting interaction effects between economic structure and agents' cognition.

However, by raising the number of goods from 3 to 4, and placing human subjects under the same conditions as our artificial agents, we were not able to replicate the results obtained by simulations. This failure may carry several interpretations. We tackle some of those thereafter. Except for the first one, they have in common to assume that an additional good greatly increases the difficulty to coordinate, which is the most probable cause of failure. (i) "It is due to specific features of the sample". We possess

data from one hundred subjects, but this corresponds to data for only two economies and we expected convergence for only one of them. It is indeed difficult to reject the possibility that the lack of convergence over a medium of exchange for the concerned economy is specific to our sample.

(ii) “The subjects (or a sub-group of the subjects) were unable to endorse the primary cost of indirect exchange (i.e., they have a strong bias towards a direct exchange strategy)”. Indeed, in a Kiyotaki & Wright environment (Kiyotaki and Wright, 1989, 1993), in the specific case where a speculative equilibrium is expected—that is to say when the monetary good has a higher storage cost than the other good—it has been noted that a non-negligible part of subjects had difficulties to endorse the primary cost implied by the use of the monetary good as a medium of exchange (i.e., to speculate) (Duffy and Ochs, 1999; Kindler et al., 2017). It means that some subjects that neither produce or consume the monetary good were reluctant to engage in indirect exchange strategies. Similarly, our experimental results show that part of the subjects that were supposed to proceed to indirect exchanges and suffer from a primary temporal cost, did not adopt such strategies, although most of the subjects that were supposed to use direct exchanges did so (see for instance the results for the condition with a non-uniform distribution promoting the good 1 with four goods depicted in the Fig. 4). As in our protocol, subjects do not play against artificial agents that use a deterministic algorithm but against other human subjects, it is nonetheless difficult to tell whether subjects playing (almost) always a direct exchange strategy did it because of the behavior of other subjects, or because they were initially strongly biased toward this option.

(iii) “Subjects were lacking information to coordinate”. Since the level of information for artificial agents was strictly identical to that of humans, it is probably for other reasons than because of a lack of information. Indeed, reinforcement learning, although effective, is far from being the most sophisticated learning model. It is unlikely that human subjects have failed to coordinate on a single medium of exchange due to more limited cognitive abilities than agents using reinforcement learning.

(iv) “The psychological model used for the simulations is unappropriated, that is the reason why it was partly ineffective in producing accurate predictions”. Several studies point out the fact that reinforcement learning models fit well the behavior of human subjects in economic contexts (Roth and Erev, 1995; Erev and Roth, 1998; Feltovich, 2000), and specifically for modeling behavior in a coordination game over a unique medium of exchange (Duffy and Ochs, 1999; Duffy, 2001; Kindler et al., 2017). However, to test the relevance of such an interpretation, we proceeded to a post hoc analysis (see Supplementary section).

We fitted the behavioral data with our reinforcement learning model, and run simulations using the best-value parameters of each subject. We obtained the same pattern as the experimental results: in the three-goods setting, the use of a medium of exchange is promoted in the condition of non-uniform distribution while in the four-goods setting, the use of a medium of exchange was not promoted as we expected. Hence, using the adequate set of cognitive parameters, we could replicate the experimental results, whether positive or null.

(v) “Assuming the cognitive model as true, this could be because the artificial agents from a single economy were having homogeneous cognitive features, while it exists certain heterogeneity among the human subjects that could make the coordination more difficult”. To test the relevance of this interpretation, after fitting the behavioral data with the model, we simulated an homogeneous population using as cognitive parameter values the average best value for each cognitive parameter after fitting the

behavioral data (instead of simulating an heterogeneous population with the parameters of a single agent being the best-value parameters of a subject fit). However, the pattern remained unchanged: the non-uniform distribution of agent types promotes the use of a medium of exchange with three goods, but not with four.

(vi) “More trials would have allowed subjects to overcome the complexity of coordination at 4 goods”. To test the relevance of this interpretation, after fitting the behavioral data with the model, we simulated a population of (heterogeneous) agents with the parameters of every single agent being the best-fit parameter values of a single subject for a larger number of iteration ($n = 500$ instead of 50). Here, the results changed (see Supplementary section), as the non-uniform distribution of agent types promotes the use of a medium of exchange in both settings with a large number of trials. This indicates that an extended time could have allowed the human subjects to modify slowly their behavior towards the use of a medium of exchange, raising the questions about the pragmatic possibility of such large scale experiments for a long time.

Nevertheless, these results seem to contribute to a better understanding of the processes underlying the coordination over a unique medium of exchange. Hence, in the 3 goods setting, the results in artificial agents, as well as those obtained in human, show that decision-makers do not need to have any expertise concerning the economic system in which they evolve to allow this system to acquire certain remarkable macroeconomic properties—such as the existence of a unique medium of exchange. Said differently, these results show that the members of an economic system do not need to know the macroeconomic properties of the system to be able to influence them.

Although, the attempt to test the robustness of the results by considering a 4 good setting appears to be unsuccessful. The results obtained by simulation and with human subjects being not completely in line, it is difficult to draw strong inferences regarding the possibility of money emergence under informational constraints in a more than three goods economy. Also, these negative results indicate the importance to take into account the temporal aspect of the coordination processes: even if we possess evidence for the existence of a steady-state for an economic system with artificial agents (or by mathematical proof), it could be that, due to the complexity of the coordination process, the time for obtaining with humans is so long that in real-world context, it would be a good approximation to say that it would never occur. At least, in the present context of money, the phenomenon already occurred, so it just remains to continue to investigate how such large scale coordination has been possible, given the complexity of the interactions.

In previous studies (Duffy and Ochs, 1999; Duffy, 2001; Lefebvre et al., 2018), subjects were constantly provided with economy statistics, such as the current distribution of goods or agent types. From this information, subjects can infer exchanges’ success probabilities. In that sense, decisions are made by description: subjects learn about the probabilistic consequences of their action by consulting descriptions of action consequences and probabilities. In contrast here, subjects are not provided with any information related to the state of the economy, decisions are therefore made by experience: subjects’ learning of outcome probabilities is based on their own experience. In the literature, one concept refers to these two kinds of decision-making systems supposed to result in behavioral discrepancies: the description-experience gap (Wulff et al., 2018).

It has been observed that decision by experience is subject to biases that are absent in decision by description. Preferential learning from positive outcomes (rather than negative outcomes)

prediction errors is for instance often observed (Palminteri et al., 2017; Lefebvre et al., 2017; den Ouden et al., 2013; Frank et al., 2007; Van Den Bos et al., 2012; Aberg et al., 2016). Interestingly, our subjects also present this asymmetry in value-update and seem to preferentially learn from exchanges that result in better-than-expected outcomes (see Fig. S18F). Investigating how such bias affects the coordination of agents in an experience-based money emergence paradigm could then constitute a relevant subject for further studies.

Data availability

The data are available at the same address than the analysis program: <https://github.com/AurelienNioche/MoneyAnalysis>.

Code availability

The software we used was based on a client/server architecture. The client part has been developed using the Unity game engine. The application ran on 7" Android tablets. The assets of the application are available at <https://github.com/AurelienNioche/MoneyApp>. The experiment server was hosted on a local server and has been developed in Python. The code of the server part is available at <https://github.com/AurelienNioche/MoneyServer>. The analysis program is available at <https://github.com/AurelienNioche/MoneyAnalysis>.

Received: 7 August 2019; Accepted: 6 November 2019;

Published online: 03 December 2019

References

- Aberg KC, Doell KC, Schwartz S (2016) Linking individual learning styles to approach-avoidance motivational traits and computational aspects of reinforcement learning. *PLoS ONE* 11(11):e0166675
- Aiyagari SR, Wallace N (1991) Existence of steady states with positive consumption in the kiyotaki-wright model. *Rev Econ Stud* 58(5):901–916
- Anbarci N, Dutu R, Feltovich N (2015) Inflation tax in the lab: a theoretical and experimental study of competitive search equilibrium with inflation. *J Econ Dynam Control* 61:17–33
- Apestequia J, Ballester MA (2018) Monotone stochastic choice models: the case of risk and time preferences. *J Polit Econ* 126(1):74–106
- Branch W, McGough B (2016) Heterogeneous beliefs and trading inefficiencies. *J Econ Theory* 163:786–818
- Brown PM (1996) Experimental evidence on money as a medium of exchange. *J Econ Dynam Control* 20(4):583–600
- Comello MLG, Qian X, Deal AM, Ribisl KM, Linnan LA, Tate DF (2016) Impact of game-inspired infographics on user engagement and information processing in an ehealth program. *J Med Internet Res* 18:9
- Davis DD, Korenok O, Norman P, Sultanum B, Wright R (2019) Playing with Money. FRB Richmond Working Paper No. 19–2. Available at SSRN: <https://ssrn.com/abstract=3333603>
- den Ouden HE, Daw ND, Fernandez G, Elshout JA, Rijpkema M, Hoogman M, Franke B, Cools R (2013) Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* 80(4):1090–1100
- Diamond P (1984) Money in search equilibrium. *Econom J Econom Soc* 1–20
- Ding S, Lugovskyy V, Puzzello D, Tucker S, Williams A (2018) Cash versus extra-credit incentives in experimental asset markets. *J Econ Behav Organization* 150:19–27
- Duffy J (2001) Learning to speculate: experiments with artificial and real agents. *J Econ Dynam Control* 25(3–4):295–319
- Duffy J, Ochs J (1999) Emergence of money as a medium of exchange: an experimental study. *Am Econ Rev* 89(4):847–877
- Duffy J, Puzzello D (2014) Gift exchange versus monetary exchange: theory and evidence. *Am Econ Rev* 104(6):1735–1776
- Erev I, Roth AE (1998) Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am Econ Rev* 88(4):848–881
- Feltovich N (2000) Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games. *Econometrica* 68(3):605–641
- Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci* 104(41):16311–16316
- Harrison GW, List JA (2004) Field experiments. *J Econ Literature* 42(4):1009–1055
- Iwai K (1996) The bootstrap theory of money: a search-theoretic foundation of monetary economics. *Struct Change Econ Dynamics* 7(4):451–477
- Jiang JH, Zhang C (2018) Competing currencies in the laboratory. *J Econ Behav Organization* 154:253–280
- Jones E, Oliphant T, Peterson P (2001) SciPy: open source scientific tools for Python
- Jones RA (1976) The origin and development of media of exchange. *J Polit Econ* 84(4, Part 1):757–775
- Kehoe TJ, Kiyotaki N, Wright R (1993) More on money as a medium of exchange. *Econ Theory* 3(2):297–314
- Kindler A, Bourgeois-Gironde S, Lefebvre G, Solomon S (2017) New leads in speculative behavior. *Phys A Stat Mech Appl* 467:365–379
- Kiyotaki N, Wright R (1989) On money as a medium of exchange. *J Polit Econ* 97(4):927–954
- Kiyotaki N, Wright R (1991) A contribution to the pure theory of money. *J Econ Theory* 53(2):215–235
- Kiyotaki N, Wright R (1993) A search-theoretic approach to monetary economics. *Am Econ Rev* 83(1):63–77
- Lefebvre G, Lebreton M, Meyniel F, Bourgeois-Gironde S, Palminteri S (2017) Behavioural and neural characterization of optimistic reinforcement learning. *Nat Hum Behav* 1(4):0067
- Lefebvre G, Nioche A, Bourgeois-Gironde S, Palminteri S (2018) Contrasting temporal difference and opportunity cost reinforcement learning in an empirical money-emergence paradigm. *Proc Natl Acad Sci* 115(49):E11446–E11454
- Luo GY (1998) The evolution of money as a medium of exchange. *J Econ Dynam Control* 23(3):415–458
- Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 18(1):50–60
- Marimon R, McGrattan E, Sargent TJ (1990) Money as a medium of exchange in an economy with artificially intelligent agents. *J Econ Dynam Control* 14(2):329–373
- Menger K (1892) On the origin of money. *Econ J* 2(6):239–255
- Muth JF (1961) Rational expectations and the theory of price movements. *Econom J Econom Soc* 29(3):315–335
- Nosal E, Rocheteau G (2011) Money, payments, and liquidity. MIT press
- Oh S (1989) A theory of a generally acceptable medium of exchange and barter. *J Monet Econ* 23(1):101–119
- Osborne M (2016) Exponential versus hyperbolic discounting: a theoretical analysis. *SSRN* 2518162.
- Palminteri S, Lefebvre G, Kilford EJ, Blakemore S-J (2017) Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLoS Comput Biol* 13(8):e1005684
- Rietz J (2019) Secondary currency acceptance: experimental evidence with a dual currency search model. *J Econ Behav Organization* 166:403–431
- Roth AE, Erev I (1995) Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. *Games Econ Behav* 8(1):164–212
- Shi S (1995) Money and prices: a model of search and bargaining. *J Econ Theory* 67(2):467–496
- Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Vol. 1. MIT Press, Cambridge
- Van Den Bos W, Cohen MX, Kahnt T, Crone EA (2012) Striatum-medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cerebral Cortex* 22(6):1247–1255
- Wanner D (2014) Serious economic games: designing a simulation game for an economic experiment. In *International conference of design, user experience, and usability*. Springer, pp 782–793
- Wright R (1995) Search, evolution, and money. *J Econ Dynam Control* 19(1–2):181–206
- Wulff DU, Mergenthaler-Canseco M, Hertwig R (2018) A meta-analytic review of two modes of learning and the description-experience gap. *Psychol Bull* 144(2):140

Acknowledgements

This work was supported by the Agence Nationale de la Recherche (ANR-16-CE38-0003). The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

Author contributions

AN and BG wrote the code, performed the experiments and the data analysis; AN, BG, GL, TB, NR, and SB-G. designed the study and co-wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1057/s41599-019-0362-2>.

Correspondence and requests for materials should be addressed to A.N.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

A.2 Interaction effects between consumer information and firms' decision rules in a duopoly: how cognitive features can impact market dynamics




ARTICLE

<https://doi.org/10.1057/s41599-019-0241-x>

OPEN

Interaction effects between consumer information and firms' decision rules in a duopoly: how cognitive features can impact market dynamics

Aurélien Nioche ^{1,2,3,4,5}, Basile Garcia^{2,3,4,5,6,7,8}, Thomas Boraud^{4,5,9}, Nicolas Rougier^{4,5,8,10} & Sacha Bourgeois-Gironde^{2,3,11}

ABSTRACT Duopolies are situations where two independent sellers compete for capturing market share. Such duopolies exist in the world economy (e.g., Boeing/Airbus, Samsung/Apple, Visa/MasterCard) and have been studied extensively in the literature using theoretical models. Among these models, the spatial model of Hotelling (1929) is certainly the most prolific and has generated subsequent literature, each work introducing some variation leading to different conclusions. However, most models assume consumers have unlimited access to information (perfect information hypothesis) and to be rational. Here, we consider a situation where consumers have limited access to information and explore how this factor influences the behavior of competing firms. We first characterized three decision-making processes followed by individual firms (maximizing one's profit, maximizing one's relative profit with respect to the competitor; or tacit collusion) using a simulated model, varying the level of information of consumers. These manipulations alternatively lead the firms to minimally or maximally differentiate their relative position. We then tested the model with human participants in the role of firms and characterized their behavior according to the model. Our results demonstrate that limited access to information by consumers can actually induce a mutually beneficial non-competitive behavior of firms, which is not traceable to explicit collusive strategies. Imperfect information on the part of consumers can hence be exploited by firms through basic and blind decision rules.

¹Aalto University, School of Electrical Engineering, Department of Communications and Networking, 02150 Espoo, Finland. ²Institut Jean Nicod, Département d'Etudes Cognitives, ENS, EHESS, PSL Research University, 75005 Paris, France. ³Institut Jean Nicod, Département d'Etudes Cognitives, CNRS, UMR 8129, 75005 Paris, France. ⁴Institut des Maladies Neurodégénératives, Université de Bordeaux, 33000 Bordeaux, France. ⁵Institut des Maladies Neurodégénératives, CNRS, UMR 5293, 33000 Bordeaux, France. ⁶Laboratoire de Neurosciences Cognitives Computationnelles, Département d'Etudes Cognitives, ENS, PSL Research University, 75005 Paris, France. ⁷Laboratoire de Neurosciences Cognitives Computationnelles, INSERM, U960, 75005 Paris, France. ⁸Inria Bordeaux Sud-Ouest, 33405 Talence, France. ⁹Centre Expert Parkinson, CHU Bordeaux, 33000 Bordeaux, France. ¹⁰LaBRI, Université de Bordeaux, INP, CNRS, UMR 5800, 33405 Talence, France. ¹¹Laboratoire d'Economie Mathématique et de Microéconomie Appliquée, Université Panthéon Assas, 75006 Paris, France. These authors contributed equally: Aurélien Nioche, Basile Garcia. These authors jointly supervised this work: Nicolas P. Rougier, Sacha Bourgeois-Gironde Correspondence and requests for materials should be addressed to A.N. (email: nioche.aurelien@gmail.com)

Introduction

Duopolies are situations where two independent sellers compete for capturing market share. There are actually numerous examples of such situations in the world economy (e.g., Boeing/Airbus, Samsung/Apple, Visa/MasterCard). What is particularly interesting in these situations is the fact that it is not rare to observe sellers adopting a similar positioning on the market, both in geographical terms and in terms of product differentiation (e.g., Burger King/McDonalds). This can appear counterintuitive at first glance: one could spontaneously assume that sellers would try to avoid such behavior in order to reduce competition. The first formal model proposed by Hotelling (1929) for describing such situations has provided an explanation on why and how firms could be incentivized to minimally differentiate. His model considers a pool of consumers that are uniformly spread over a one-dimensional segment. Two firms selling the same product have to decide where to locate on this segment and what price to offer for their product, knowing that each consumer will choose a firm according to its relative distance (linear transportation costs) and the price of the product. The original study holds that in such conditions, firms tend to aggregate and compete near the center of the segment (minimal differentiation principle) due to the effort of the firms to capture the largest number of consumers. However, subsequent research (d'Aspremont et al., 1979; Cremer et al., 1991; Economides, 1993; Brenner, 2005) demonstrated the existence of an antagonist principle of maximal differentiation, using either quadratic transportation costs, a higher number of competitors or a higher number of dimensions on which firms can differentiate themselves. In the end, both minimal and maximal differentiation can be incentivized and observed (Irmen and Jacques-François, 1998). Here, we show how the level of information of consumers may induce different behavior for the two firms, depending on their strategies.

Several experimental studies have already attempted to characterize the various factors influencing differentiation. For instance, Kruse et al. (2000) allowed for communication between participants in the role of firms and showed that they tend to group in the center when communication is limited, but on the contrary, to differentiate themselves if communication is unlimited (cooperation). Similarly, Kephart and Friedman (2015) set-up a protocol contrasting continuous and discrete time and demonstrated that continuous time could trigger a maximal differentiation strategy, as it allows some form of communication, and as a consequence, some form of cooperation. These two findings brought together suggest that quick and/or full information transmission can help to reach a cooperative equilibrium in a typical Hotelling's model. Several other studies brought up arguments supporting the robustness of the minimal differentiation phenomenon such as, for example, the four-player version of the game by Huck et al. (2002) or in Barreda-Tarrazona et al. (2011), where subjects tended to group in the center under several experimental conditions. Although there is a treatment in Barreda-Tarrazona et al. (2011) with human subjects as consumers, what is common to all these works is their shared assumption of the fact that firms are competing to capture rational and fully informed consumers even when they document spatial behavior that departs from Nash equilibrium when it theoretically exists. The case when consumers have no full informational access to the firms' strategies and when firms must compete over this less than completely informed consumers have not been addressed, to our knowledge, in the experimental literature related to Hotelling (1929). It has yet important implications as it is a common fact that consumers are not fully aware of all options available on the markets they participate in and that firms know and anticipate this fact in their own strategies.

Stigler (1961) argued that the information question should be fully taken into account in such competition models, as it can deeply impact the nature of equilibria. This is particularly important as consumer choices are known to be subject to several biases and based on partial information (Thaler, 1980; Kahneman, 2003). More precisely, the uncertainty resulting from the imperfect nature of information has been shown to provide an incentive for the firms to regroup and transparency of the market is, thus, a prominent factor for differentiation (Webber, 1972; Stahl, 1982). In line with predictions from earlier studies (Eaton and Richard, 1975; Brown, 1989; Dudey, 1990; Schultz, 2009), we postulate that in a duopoly context, the consumers' access to information is a critical factor for the differentiation of the two firms.

We thus defined a formal turn-based model (Prescott and Vissher, 1977; Loertscher and Muehlheusser, 2011) that allows us to explicitly manipulate the amount of information available to consumers while retaining their rational nature. Agents can act rationally under partial information and thereby induce observable organizational patterns in the market that differ from what is expected under perfect information. We test the hypothesis that the amount of information accessible to consumers can variably drive the differentiation of the two firms: when this amount is low, firms will be maximally differentiated; when this amount is high, firms will be minimally differentiated. We test this hypothesis using a simulation where we consider three decision-making processes for the firms, namely (i) a maximization of short-term profits, (ii) a maximization of the difference of profits between the firm and its opponent, (iii) a maximization of the profits of the two firms. Following Rubinstein's prescriptions (Rubinstein, 1991), the aim of these decision rules is to incorporate the potential perception of the situation by the decision makers. These rules indeed constitute plausible behavioral responses on the part of firms in the light of partial information on the part of consumers. These decision rules helped us to characterize the behavior of human subjects for the experimental part of this work where subjects play the role of the firms under different informational conditions. The choice of using the first decision rule is straightforward: a firm only pays attention to its current own profit, considering further expectations about the future not reliable. A firm may consider the behavior the other firm either too difficult to compute or not reliable. This is tantamount to ignore the other player strategies. This type of "blind" decision rule has a presence in the Industrial Organization literature stemming as far as Rothschild (1947) in which securing one's profit is the only rule by which a firm's behavior is guided.

Regarding the second decision rule, motivations are dual. It could lie on an anchoring bias (Tversky and Kahneman, 1974): it is difficult to evaluate the success of a move per se, a move is considered efficient if it leads to better profits than its opponent. In other words, firms' strategies evaluation relies on comparisons to a given point instead of evaluation in absolute terms. Secondly, it could be due to a zero-sum bias (Meegan, 2010; Rózycka-Tran et al.), even if in our model, the profit of one firm is not necessarily made at the expense of the other firm (see 'Methods' section). Indeed, considering—sometimes wrongfully—that a greater profit for its opponent is a profit loss for itself, a firm could decide to make its choice only considering the profit difference.

The use of the third heuristic lies on the expectation of tacit collusion (TC) with the opponent: if both firms try to maximize their own profit as well as the profit of their opponent, it would avoid the drawbacks of a competitive situation and leads to higher profits. It could be also seen as a search for Pareto optimality (Pareto, 1964), that is to say following the strategies that lead to a

repartition of profits such as no firm could earn more, otherwise, it would be at the expense of the other.

Therefore, we hypothesized that (i) depending on the information available to consumers, either a minimum differentiation or a maximal differentiation can apply, (ii) the effect of the information available to consumers can be modulated by the firms' decision rules.

Methods

Model description. We consider a unidimensional normalized space X discretized into n_{cons} evenly spread locations such that $x_i = (i - 1)/(n_{\text{cons}} - 1)$. We consider a set of $1 + p_{\text{max}} - p_{\text{min}}$ integer prices P ranging from p_{min} to p_{max} . We consider two firms $\{F_j\}_{j \in \{1,2\}}$ and a group of n_{cons} consumers $\{C_i\}_{i \in \{1, \dots, n_{\text{cons}}\}}$.

Each firm $F_i = (x_i, p_i)$ is characterized by a position x_i and a price p_i . Consumers are uniformly spread over space such that $x_i = (i - 1)/(n_{\text{cons}} - 1)$. View radius is defined on a per-experiment basis and is the same for all the consumers during an experiment. The firm position is a free variable and must correspond to a consumer position such that there are only n_{cons} different possible positions for a firm. Price p_i is a free variable and is discrete: there are P possible prices spread uniformly between a minimal price p_{min} and a maximal price p_{max} . Simulations are turn-based (Prescott and Vissher, 1977; Loertscher and Muehlheusser, 2011). We distinguish at each turn an active firm that is allowed to select a strategy and a passive firm that has to wait for the next turn to react and deploy its own strategy. More specifically, at turn t , Firm A (F_1 or F_2) chooses its location and its price, consumers choose a firm and profits are collected for both firms. At turn $t + 1$, Firm B (F_2 or F_1) chooses its location and its price (while Firm A keeps location and price from turn t), consumers choose a firm and profits are collected for both firms. Let Π_i be the profit of the firm F_i for a single turn defined by $\Pi_i = p_i \cdot q_i$ with p_i the price at which F_i sells its product, and q_i the quantity F_i sold. There is no production cost. Consumers are able to buy only one product per turn. They consume it immediately, in such a manner that they do not constitute any stock. This implies that a firm produces a maximum of n_{cons} products per turn.

Each consumer $C_i = (x_i, r_i)$ is characterized by a position x_i and a view radius r_i . The view radius defines a segment centered on the consumer $[x_i - r_i, x_i + r_i]$. Only firms located inside this segment are considered by the consumer (see Fig. 1). Consequently, at each turn, some consumers will see only one firm and will be captive since they cannot choose what firm to buy from. Some consumers will see both firms and are named volatile because they can choose any of the two firms depending on their choice criterion. Some consumers won't see any firms and cannot buy, and, thus, are named ghost consumers. A view radius of 0 means the firm has to be at the same position to be seen while a radius of 1 means the firm is seen by all the consumers. Reciprocally, and depending on the consumer view radius, firms have access only to a subset of all the consumers, they are named the potential consumers and represent the sum of captive and volatile consumers.

Parameters. For all the simulations, we used the following parameters: n_{cons} (number of consumers) = 21, n_{price} (number of prices) = 11, p_{min} (minimal price) = 1, p_{max} (maximal price) = 10, $n_{\text{turn}} = 100$. The initial position and price for the passive firm (first turn) are randomly assigned. The view radius (r) is the same for all the consumers and is comprised between 0 and 1. For each of the three different decision-making processes, we ran (i) 1000 simulations with r randomly (uniformly) drawn between 0 and 1 for each simulation; (ii) 64 additional

simulations with $r = 0.25$ and $r = 0.50$ respectively, in order to characterize experimental data.

Decision-making processes. Consumers do not choose the amount of information they dispose of. They may see zero, one or two firms. In the event that they do not see any firm, they are unable to buy and have to wait for the next turn. If they see a single firm, they have no means to compare prices and have to buy from this firm, independently of the price (each consumer has an unlimited budget). When they are able to see the two firms, they buy from the firm offering the lowest price. In the specific case where prices of both firms are equal, they choose randomly between the two. Firms have perfect knowledge of the environment: they know (i) the price and the position of the opponent, (ii) the location of each consumer x_i , (iii) the view radius r_i of each consumer and (iv) the decision-making method of consumers. Firms from two different simulations can differ in their decision-making process but two firms from the same simulation share the same decision-making process. Decision-making process of a firm is either one of the three following decision rules:

Profit maximization (PM). Each time an active firm plays, it computes the potential profits for all the possible position-price couples regarding the current decision-making process of the passive firm, and chooses the couple position-price that maximizes profit (in the case where several couples position-price lead to the same best payoff, the firm chooses randomly between these moves);

Difference maximization (DM). If Firm A is the active firm, the difference between its own profit and the profit of Firm B is computed for each possible move, and the move leading to the greatest difference is chosen (in case of multiple moves leading to the greatest difference, the move is randomly chosen between those moves);

Tacit collusion (TC). The distance to the maximum profit for Firm A and the distance to the maximum profit for Firm B are computed for each move the active firm can play, the chosen move is the one leading to the minimum sum of the distances.

Firm decision-making process. The choice space is defined by the set Y , that is the Cartesian product of all possible location-price couples:

$$Y = \left\{ \left(x_i, p_j \right) \right\}_{x_i \in X, p_j \in P}$$

The expected profit of the firm A is defined by the number of potential consumers (the sum of captive and volatile consumers) that a firm can expect when making the move y_α , that is locating at position x_i and setting price p_j .

Let the boolean-valued function V_{C_k} determine if a consumer C_k sees the location x_i :

$$V_{C_k}(x_i) = \begin{cases} 1 & \text{if } x_k - r_k \leq x_i \leq x_k + r_k, \\ 0 & \text{otherwise.} \end{cases}$$

Let the function V_{C_k} define the profit that the firm A can expect from a consumer C_k for the move $y_\alpha = (x_i, p_j)$, knowing that the

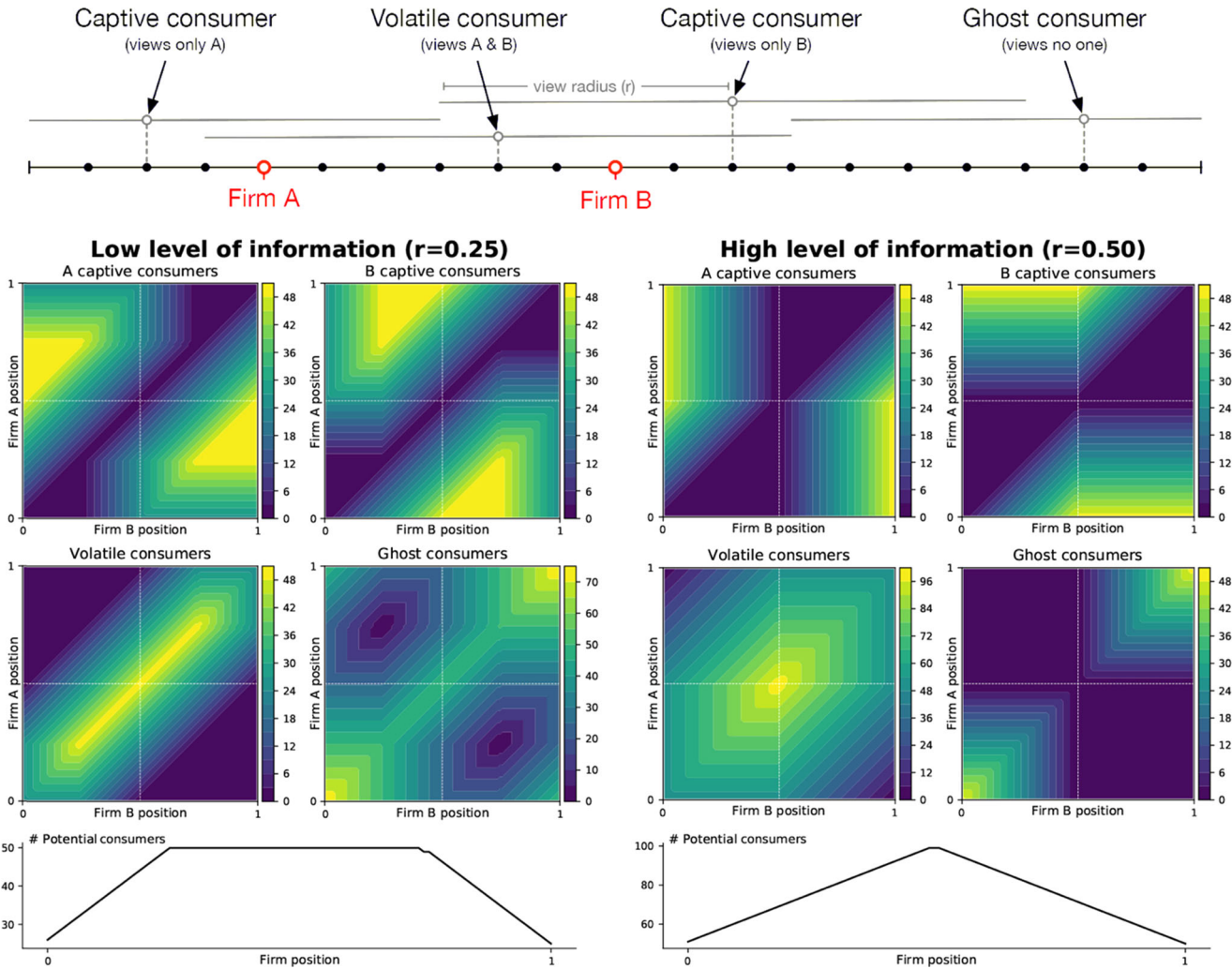


Fig. 1 Model. The model is represented by a one-dimensional line segment over which consumers (black dots) are spread uniformly. Firms (outlined red dots) are free to position themselves on any consumer position. Consumers view firms that fall within their view radius such that some consumers view only one firm (captive consumers), some consumers view both firms (volatile consumers) and some consumers view none (ghost consumers). The number of captive, volatile, and ghost consumers for a firm depends on the size of the view radius of the consumers and the respective position of the two firms. The number of captive consumers increases as the view radius of consumers decreases. The number of potential consumers (captive + volatile) of a firm is a function of the position of the firm and the view radius of the consumers. When the level of information is high ($r = 0.50$), there is a unique position where all the consumers are potential consumers for a firm. When the level of information is low ($r = 0.25$), there is a whole segment where exactly half the consumers are potential consumers for a firm

move of the opponent is $y_\beta = (x_m, p_n)$:

$$W_{C_k}^A(y_\alpha, y_\beta) = \begin{cases} 0 & \text{if } [p_j > p_n \wedge V_{C_k}(x_m) = 1] \vee V_{C_k}(x_i) = 0, \\ 0.5 & \text{if } p_j = p_n \wedge V_{C_k}(x_m) = 1, \\ 1 & \text{otherwise.} \end{cases}$$

Then, the expected profit of the firm A for a given move y_α is obtained from the function E^A , knowing that the move of firm B is y_β :

$$E^A(y_\alpha, y_\beta) = \sum_{k=1}^{n_{\text{cons}}} W_{ck}^A(y_\alpha, y_\beta)$$

A firm makes the move y_λ (that is a specific combination of position and price) using one of the following decision rules: profit maximization (PM), difference maximization (DM), or tacit collusion (TC).

When following a PM decision rule, the active firm A maximizes its expected profit $E^A_{y_i, y_\beta}$ such that:

$$\lambda = \arg \max_i \left(\left\{ E^A(y_i, y_\beta) \right\}_{y_i \in Y} \right)$$

When following a DM decision rule, the active firm A maximizes the difference between its own expected profit and the corresponding expected profit of the passive firm B:

$$\lambda = \arg \max_i \left(\left\{ E^A(y_i, y_\beta) - E^B(y_i, y_\beta) \right\}_{y_i \in Y} \right)$$

When following a TC decision rule, the firm A makes the move maximizing both its own expected profit and the expected profit of its opponent, by considering the relative distance to the

maximum expected profit for both firms:

$$\Delta^A(y_\alpha, y_\beta) = \max_{y_i \in Y} \left(\left\{ E^A(y_i, y_\beta) \right\} \right) - E^A(y_\alpha, y_\beta)$$

$$\lambda = \arg \min_i \left(\left\{ \Delta^A(y_i, y_\beta) + \Delta^B(y_i, y_\beta) \right\}_{y_i \in Y} \right)$$

Experiments. Participants were recruited using the Amazon Mechanical Turk (AMT) platform. AMT is an online crowdsourcing service where anonymous online workers complete web-based tasks in exchange for monetary compensation. It was noted that responses from AMT participants were at least as reliable as those obtained in laboratories (Buhrmester et al., 2011; Amir et al., 2012). In addition, AMT participants exhibit similar judgment and decision biases such as framing effects, conjunction fallacy, or outcome bias (Paolacci et al., 2010). The ethics approval for this project was provided by the Ecole Normale Supérieure as per the school's guidelines. In line with ethical guidelines, all participants provided informed consent before proceeding to the experiment. Participants also had to fill in a survey asking their age, nationality, and gender. Monetary compensation of one dollar was offered to each participant, with a bonus proportional to their score. In average, participants received a compensation of \$2.64 (± 0.58 SD). Participants were paired inside a dedicated virtual room and each pair went through one of the four treatments. The four treatments correspond to the combination of two factors: consumers' view radius (r) and the display of the opponent's profit (s). The consumers' view radius that was either low ($r = 0.25$) or high ($r = 0.50$) and the opponent's profit was either hidden ($s = 0$) or displayed ($s = 1$). For all the rounds, we used the same parameters as for simulations except that we maintained constant the initial locations of firms: one of the two firms was placed at one of the extrema of the segment, the other firm at the other extrema. Their initial price was set to 5. The subject playing first was randomly selected. The number of rooms (with two subjects each) for each condition is: ($r = 0.25, s = 1, n_{eco} = 26$), ($r = 0.25, s = 1, n_{eco} = 30$), ($r = 0.50, s = 1, n_{eco} = 26$), ($r = 0.50, s = 1, n_{eco} = 26$). Additional information is provided in the supplementary section.

Analysis. Only data obtained from subjects that have fully completed the experimental procedure has been used for analysis. Among the 410 subjects that signed up to the platform, 222 subjects went through all the process (see Supplementary for more information). The reasons why a subject may not have completed the procedure are (i) the impossibility to match him with another subject, (ii) quit before the end, (iii) a technical problem (i.e., poor computer performances). The sample of subjects we used for analysis matches the demographic characteristics of AMT (Ipeirotis, 2010). Regarding the composition of the participants, we noticed a quasi-gender parity (women represented 54.1% and men 45.9%). The average age was 34.75 ± 9.54 . We counted a dozen nationalities, the most common being American (75.78%) with a large majority, and, to a lesser extent, Indian (11.71%).

We drew a three-dimensional profile for each subject. Each dimension corresponds to a particular decision rule (PM, DM, TC). The score associated with each dimension assesses the extent to which a subject behaves accordingly to what a specific decision rule implies to do.

Let be $v_H^A(y_\alpha, y_\beta)$ be the value of the move y_α relatively to the decision rule $H \in \{PM, DM, TC\}$ (respectively, for PM, DM,

TC) at time $t \in [1, n_{turn}]$:

$$v_{PM}^A(y_\alpha, y_\beta) = \begin{cases} \frac{E^A(y_\alpha, y_\beta)}{\max(\{E^A(y_i, y_\beta)\}_{y_i \in Y})} & \text{if } \max(\{E^A(y_i, y_\beta)\}_{y_i \in Y}) > 0, \\ 1 & \text{otherwise.} \end{cases}$$

$$v_{DM}^A(y_\alpha, y_\beta) = \begin{cases} \frac{E^A(y_\alpha, y_\beta) - E^B(y_\alpha, y_\beta)}{\max(\{E^A(y_i, y_\beta) - E^B(y_i, y_\beta)\}_{y_i \in Y})} & \text{if } \max(\{E^A(y_i, y_\beta) - E^B(y_i, y_\beta)\}_{y_i \in Y}) > 0, \\ 1 & \text{otherwise.} \end{cases}$$

$$v_{TC}^A(y_\alpha, y_\beta) = \begin{cases} \frac{\min(\{\Delta^A(y_i, y_\beta) + \Delta^B(y_i, y_\beta)\}_{y_i \in Y})}{\Delta^A(y_\alpha, y_\beta) + \Delta^B(y_\alpha, y_\beta)} & \text{if } \Delta^A(y_\alpha, y_\beta) + \Delta^B(y_\alpha, y_\beta) > 0, \\ 1 & \text{otherwise.} \end{cases}$$

For convenience, we did not include the variable t in the definition of the v functions. So, let's assume a function f such as:

$$f_H(i, t) = v_H^i(y_\alpha, y_\beta) \text{ for time } t$$

The score for a subject i and for a decision rule t is simply the average value of f over time:

$$s_H(i) = \frac{1}{n_{turn}} \sum_t^{n_{turn}} f_H(i, t)$$

For the analysis, we pooled the individual scores by experimental condition. As we did not expect a normal distribution of the data due to clustering effects at the boundaries of our scales (i.e., price), assessment of statistic relevancy of our observations has been made with Mann-Whitney's U -ranking test, applying Bonferroni's corrections for multiple comparisons. We set the significance threshold at 1%.

Results

Simulations. In order to test our hypothesis regarding the influence of the information level of consumer (measured by his view radius) on the differentiation of the two firms, we ran 1000 simulations using a random view radius between 0 and 1 and tested three different decision rules for the firms, namely PM, DM, and TC. For each of these simulations, we measured the mean distance between the two firms, which is the distance separating the two firms averaged over the last third of the 100 turns (i.e., the last 33 turns). We report in Fig. 2 all these distances on the y -axis and the corresponding view radius on the x -axis. Minimal differentiation corresponds to a mean distance of 0 firms being placed at the center of the linear city while maximal differentiation corresponds to a mean distance of 0.5 one firm being placed on the first quarter of the linear city and the other one at the last quarter.

The high dispersion of the points when r is close to 0 or near 1 can be explained by the fact that the firm location has almost no impact on the firm profits. If the value of r is close to zero, the consumers are almost blind in the sense that their view radius is so narrow that except if the competitor is very close, each firm would sell its product to only a few consumers, regardless of its position. If r is close to 1, the visual field of the consumer is so broad that it will see both firms and these firms will always compete. For such extreme values of r , the mean distance

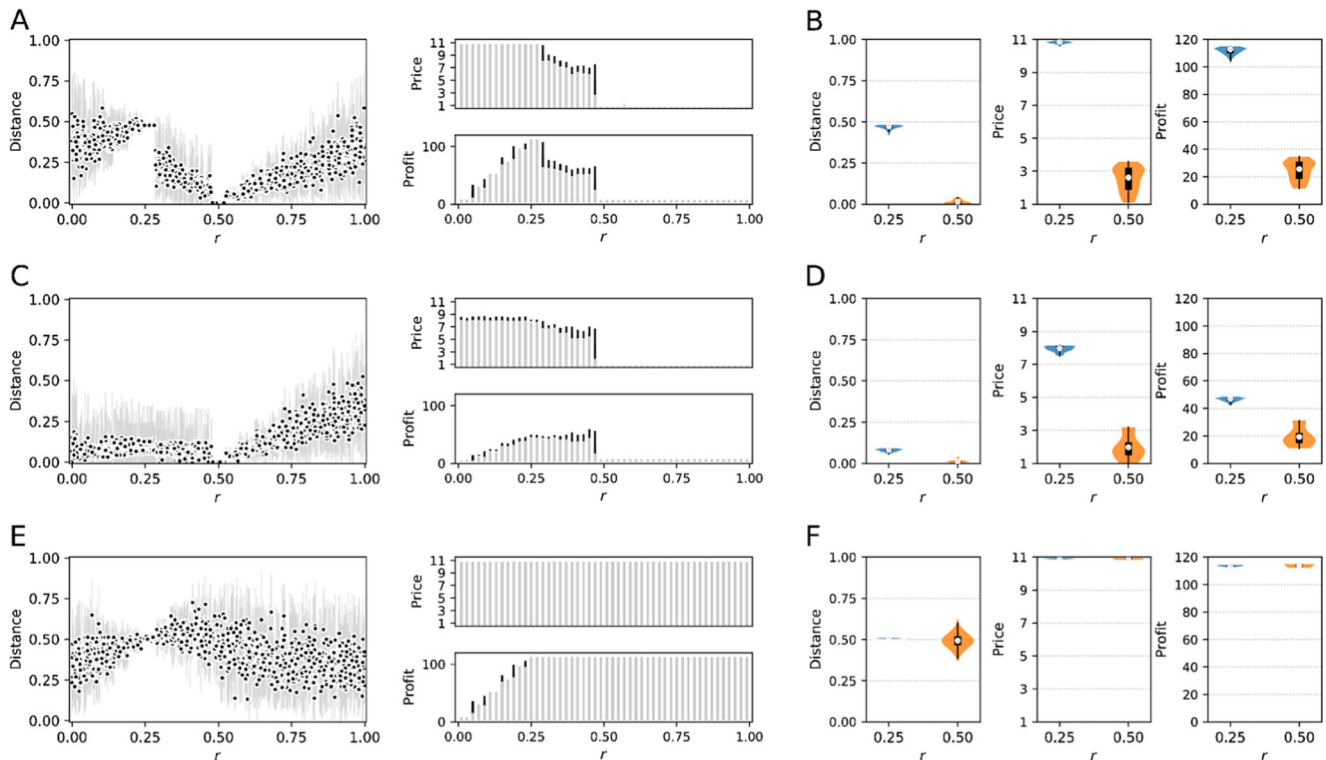


Fig. 2 Simulation results. For each of the three decision rules (profit maximization [PM], difference maximization [DM], tacit collusion [TC]), 1000 simulations were run with a random (uniform) view radius for the consumers. The distance between the two firms, the profit and the price as a function of the consumers' view radius is displayed in **a** (PM), **c** (DM), and **e** (TC). Each dot corresponds to the mean distance that has been observed between the two firms during a single simulation and vertical bars indicate the standard deviation. Mean prices and profits are reported on the right using gray bars and the standard deviation in black. For each of the three decision rules (PM, DM, TC) and for low- and high-view radius ($r = 0.25$, $r = 0.50$), 25 additional simulations were run and observed distance, price, and profit are displayed on the right (**b**, **d**, **f**) in order to compare them to the experimental results

observed is close to 0.33, which corresponds to the mean distance observed for random moves. As each consumer sees only its own position or sees both firms, it is indeed expected that the firms randomly choose their location, which corresponds to a mean distance of 0.33.

We can observe in Fig. 2a that for the PM decision rule, a view radius of 0.50 corresponds to the minimal differentiation principle where the two firms compete to occupy the central position because this is the unique position that gives access to all the consumers (all consumers are potential consumers for the firm positioned at the center). It thus makes sense for the two firms to compete around this position and to try to get a maximum number of consumers in order to maximize their profit. Because of this competition, the mean price for both firms is very low and leads to moderate profits. When the view radius is reduced to 0.25, the mean distance between the two firms is maximal (0.5). This specific radius corresponds to a case where there is a possibility of local markets as shown on Fig. 1. The two ends of the plateau when $r = 0.25$ represent a compromise between competition and a lesser number of consumers, but those consumers are captive for each firm. This allows both firms to set higher prices and to maximize their profits. When the DM decision rule is used (see Fig. 2c), firms tend to minimally differentiate, the proximity forcing them to reduce their prices and hence, greatly reducing their profits compared to what happens with firms using a PM decision rule. As one would expect, when firms try to optimize at the same time their profit and their opponent's profit (TC decision rule), firms tend to maximally differentiate and establish local monopolies (see Fig. 2e).

Focusing our attention on the specific cases where $r = 0.25$ or $r = 0.50$ (Fig. 2b, d and f), one can notice quite different situations in terms of distance, price and profits for the three policies respectively. For $r = 0.25$, the principle of maximal differentiation applies for PM and TC decision rules, leading to maximal prices and profits. This is not true for the DM decision rule where the principle of minimal differentiation seems to apply, leading to moderate prices and profits. For $r = 0.50$, the situation is different and both the PM and DM decision rules lead to a minimal differentiation of the two firms with low prices and profits. Only the tacit collusion decision rule (TC) allows for an implicit equal share of the market, with highest prices and profits. Together, these three decision rules allow to give account on minimum or maximum differentiation in the two specific cases of low and high level of information available to the consumers.

Experiments. When considering the effect of the view radius of consumers on the mean distances, prices and profits (each of these measures being an average for each pair of subjects, in such a manner that each observation accounts for a two subjects couple), experimental results are very similar to the results of simulations when the PM decision rule is used, and this, independently of whether the opponent's profit is visible or not. Indeed, a large view radius induces a minimal differentiation effect where the two firms are led to a fierce competition around the central location, subsequently decreasing their prices and profits. Conversely, when consumers dispose of a narrow view radius, firms tend to exploit this disability by locating at the endpoints of the segment.

More precisely, the median distance is greater when $r = 0.25$ than when $r = 0.50$ (when $s = 0$, $u = 81.5$, $p < 0.001$, $n = 59$; when $s = 1$, $u = 33.0$, $p < 0.001$, $n = 52$). The same applies for the prices (when $s = 0$, $u = 128.5$, $p < 0.001$, $n = 59$; when $s = 1$, $u = 30.5$, $p < 0.001$, $n = 52$) and for the profits (when $s = 0$, $u = 178.0$, $p < 0.001$, $n = 59$; when $s = 1$, $u = 131.5$, $p < 0.001$, $n = 52$).

The display of the opponent's profit ($s = 1$) has a limited effect on the general shape of the data relatively to distance, price, and profit. It has an effect on distance only when $r = 0.25$ (when $r = 0.25$, $u = 191.0$, $u = 191.0$, $p = 0.007$, $n = 56$; when $r = 0.50$, $u = 283.0$, $p = 0.686$, $n = 55$) and no effect on price (when $r = 0.25$, $u = 321.0$, $p = 1.000$, $n = 56$; when $r = 0.50$, $u = 311.0$, $p = 1.000$, $n = 55$) and profit (when $r = 0.25$, $u = 252.0$, $p = 0.143$, $n = 56$; when $r = 0.50$, $u = 296.5$, $p = 1.000$, $n = 55$).

A table summarizing the results is available in the supplementary section (see Table S4).

Although the general shape of data is close to what has been observed with simulations using the PM decision rule, the dispersion of results is much more spread out and we assume this scattering of the data can be attributed to inter-individual differences. In order to study this inter-individual variability, we computed three individual scores for each subject, assessing the compatibility of their behavior for each time step of the experiment with the use of (i) a PM decision rule, (ii) a DM decision rule, (iii) a TC decision rule. Distribution by experimental condition of PM, DM, and TC scores are shown in Fig. 3b. A matrix correlation of the scores by experimental condition has also been computed (see Fig. 3c).

Considering the effect of the field of view on individual scoring, the results indicate that the DM scores are higher in condition of high information while PM scores are lower. The variation of the view radius has no significant impact on TC scores. More precisely, considering the effect of field of view on individual scoring and comparing when $r = 0.50$ to when $r = 0.25$, for both value of s , we observe that the DM score are significantly higher (when $s = 0$, $u = 488.0$, $p < 0.001$, $n = 118$; when $s = 1$, $u = 460.0$, $p < 0.001$, $n = 104$) and the PM scores are significantly lower (when $s = 0$, $u = 1048.0$, $p < 0.001$, $n = 118$; when $s = 1$, $u = 861.0$, $p < 0.001$, $n = 104$). Still when $r = 0.50$ compared to when $r = 0.25$, the TC scores are significantly lower, but only when $s = 1$ (when $s = 0$, $u = 1452.5$, $p = 0.730$, $n = 118$; when $s = 1$, $u = 461.5$, $p < 0.001$, $n = 104$).

Considering the effect of the display of the opponent's score, the results indicate that if the opponent's score is displayed, the DM scores are higher in condition of low information. In contrary, it has no significant impact on PM and TC scoring. More precisely, considering the effect of opponent's profit displaying on individual scoring (i.e., when $s = 1$ compared to when $s = 0$), the DM score are significantly higher only when $r = 0.25$ (when $r = 0.25$, $u = 733.0$, $p < 0.001$, $n = 112$; when $r = 0.50$, $u = 1031.0$, $p = 0.026$, $n = 110$), while the PM scores are not statistically different (when $r = 0.25$, $u = 1248.5$, $p = 0.413$, $n = 112$; when $r = 0.50$, $u = 1438.5$, $p = 1.000$, $n = 110$), neither are the TC scores (when $r = 0.25$, $u = 1204.5$, $p = 0.227$, $n = 112$; when $r = 0.50$, $u = 1376.5$, $p = 0.216$, $n = 110$). A table summarizing the results is available in the supplementary section (see Table S5).

Also, similarly to the results obtained by simulation, a radius value of 0.25 allows to discriminate the use of a PM decision rule from a TC decision rule, and a radius value of 0.50 allows to discriminate the use of a PM decision rule from a DM decision rule. This is especially noticeable when looking at the distribution of the scores (Fig. 3b) but also when looking at the correlation matrix (Fig. 3c). When $r = 0.25$, a subject who has a high score in PM would likely to have a high score in TC but a low score in DM, while when $r = 0.50$, a subject who has a high score in PM

would likely have a high score in DM but a low score in TC. Hence, when trying to discriminate different decision rules, a condition of low information ($r = 0.25$) allows to distinguish a PM from a DM decision rule, but not from a TC decision rule. Conversely, in a condition of high information ($r = 0.50$), a DM is indistinguishable from a PM decision rule, but a TC decision rule is. In other words, a DM decision rule could be interpreted as a decision rule revealed solely in a condition of low information, while a TC decision rule could be interpreted as an adaptive decision rule in condition of high information.

If we now look more closely at individual behaviors, it is striking to see that when subjects competing together have been identified both as users of a specific decision rule (i.e., obtained a high score toward PM, DM, or TC), the dynamics of their playing is very similar to the corresponding simulation. With $r = 0.25$, subjects using PM decision rule tend to position themselves at the first and third quarters of the segment and both set a high selling price (see Fig. 4a). However, when $r = 0.50$, subjects position themselves at the center and immediately lower their price (see Fig. 4b) even though they are less inclined to do so compared to simulated firms using the corresponding decision rule. They are actually trying to regularly increase their price. The situation is a bit different for subjects using a DM decision rule when $r = 0.25$ (see Fig. 4c). In that case, the positions of subjects oscillate around the center accompanied with an increase and decrease in prices, indicating a will to capture the market of their opponent. When $r = 0.50$, both simulated firms and human subjects using a TC decision rule set their prices at their maximum but the dynamics are not exactly the same (see Fig. 4f). Subjects positioned themselves further apart, and this increase of the distance can be assumed to be due to an intent from the subjects to communicate their goodwill to their opponent.

Discussion

The principle of minimal differentiation as exposed in the seminal paper of Hotelling (1929) did not reach consensus in the abundant subsequent literature. Once some restrictive assumptions of the initial model are relaxed (for instance, number of firms, spatial structure, or cost structure), it has been shown that the principle of minimal differentiation can be invalidated and that the antagonistic principle of maximal differentiation could apply (d'Aspremont et al., 1979; Cremer et al., 1991; Economides, 1993; Brenner, 2005). In addition of these theoretical results, several experimental studies show that by manipulating either the communication between firms (Kruse et al., 2000) or by manipulating the time structure (Kephart and Friedman, 2015)—what also indirectly impacted the ability of the firms to communicate about their intentions—it was possible to induce either a minimal or a maximal differentiation between the firms. Similarly to Kruse et al. (2000) and Kephart and Friedman (2015), and despite the robustness of the minimal differentiation principle highlighted by the experimental results of Huck et al. (2002) and Barreda-Tarrazona et al. (2011), our results report both phenomena: simulations and experiments allowed us to demonstrate that the consumers' amount of information affects the differentiation of firms with respect to their decision-making strategies. We isolated incentives supporting either a geographic concentration and a fierce price competition resulting in drastic reduction of profits, or a maximal differentiation inducing a softening of the price competition and thereby a large increase in firms' profits. However, our results also show that the principle of maximal differentiation may be systemic and cannot be uniquely attributed to the deliberate use of a cooperative strategy on the part of firms (as in Kruse et al., 2000), or to TC (as in Kephart and Friedman, 2015).

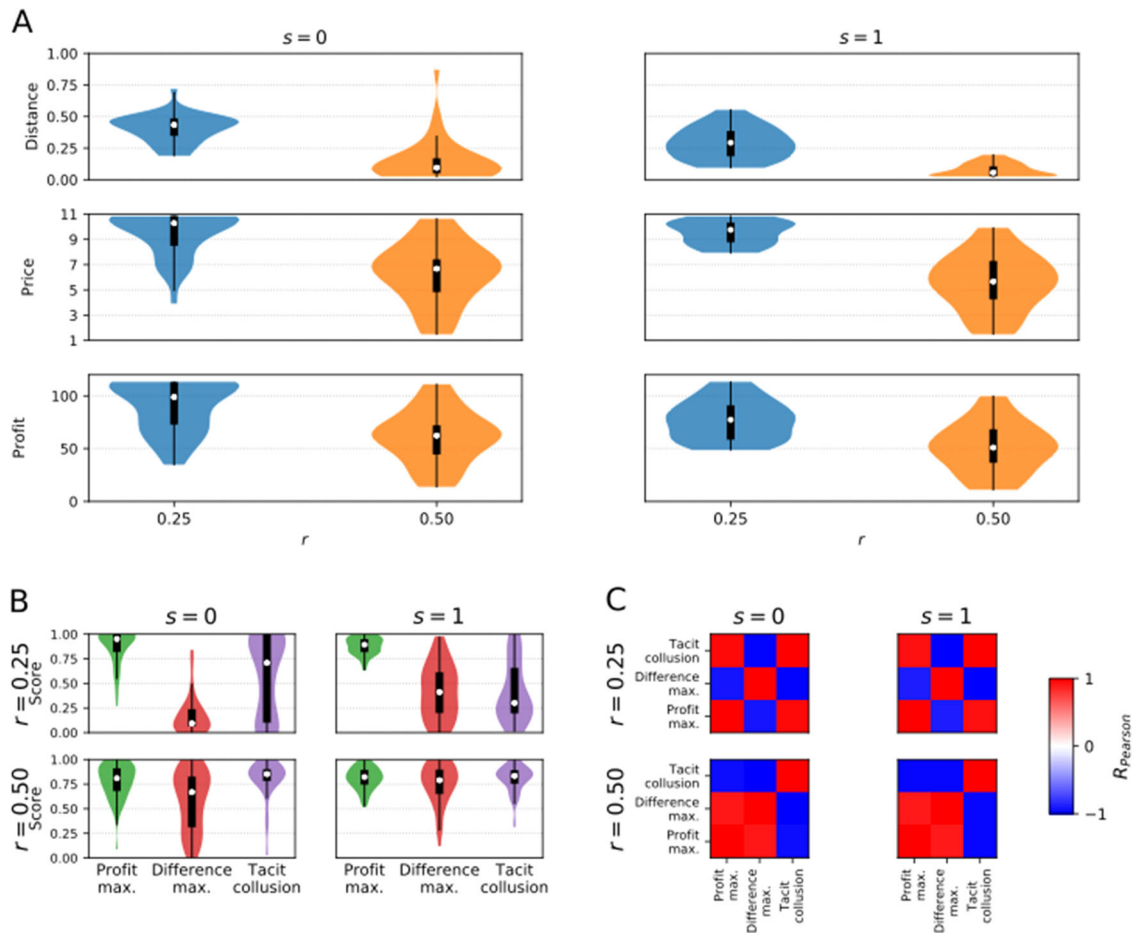


Fig. 3 Experimental results. **a** Combined effect of the consumers' view radius and the display of the opponent's score on distance, price, and profit. The white dots indicate the median, the thick black bars indicate the IQR. The extrema of the thin bars indicate the lower and upper adjacent values. The colored areas give an indication of the shape of the data distribution. **b** Mean scores of profit maximization (PM), difference maximization (DM), and tacit collusion (TC) by experimental condition. The white dots indicate the median, the thick black bars indicate the IQR. The extrema of the thin bars indicate the lower and upper adjacent values. The colored areas give an indication of the shape of the data distribution. **c** Score correlation matrix. Blue color indicates a strong negative correlation and red color a strong positive correlation

Indeed, when consumers have only access to a low level of information, the occurrence of maximal differentiation in experimental results can in turn be interpreted as an adaptation to these consumers' limited access to information. In that circumstance, firms using a PM decision rule formed local monopolies without any willingness to cooperate with the other firm. This supports and provides a possible rationale to d'Aspremont et al.'s final open remark in their fundamental reexamination of Hotelling's model (1929), according to which, contra Hotelling, one should intuitively expect differentiation to be a distinctive feature of oligopolistic competition. Oligopolists should indeed be better off by dividing the markets into submarkets over which they each exert quasi-monopolistic control. Our results actually demonstrate that limited access to information, on the part of consumers, can be an underlying factor and a prevailing one in actual competitive markets that induces a non-competitive behavior from which firms, without prior explicit collusion, can take advantage of the situation and establish local monopolies, which are detrimental to consumers. Our results show that only the use of a profit DM decision rule precludes the formation of such local monopolies.

Besides, the use of these decision rules allowed us to emphasize heterogeneous behaviors. Thinking of these various behaviors in terms of deviation from a rational behavior understood as the maximization of a unique utility function would have prevented

us from making sense of this heterogeneity. In order to define our decision rules, we measured whether our subjects looked for maximizing their own profit, whether they aimed at maximizing the difference of profits with their opponents, or finally whether they tried to create a TC. The PM decision rule appears to be a good predictor of the firms' aggregated behavior, while the other decision rules offer an opportunity to account for less expected behaviors.

The use of a DM decision rule indeed supported a fierce competition when informational structure opened the possibility of quasi-monopolies. The use of this decision rule could be explained by the presence of an underlying anchoring bias (Tversky and Kahneman, 1974): as it is difficult to evaluate the success of a move per se, a move is considered efficient if it leads to beer profits than its opponent. In other words, firms' strategy evaluation relies on comparisons to a given point instead of an evaluation in absolute terms. This could explain why this decision rule has been promoted by the display of the opponent score. The use of such decision rule could also be due to an underlying zero-sum bias (Meegan, 2010; Rózycka-Tran et al.): considering wrongfully that a greater profit for its opponent is necessarily a profit loss for itself, a firm could decide to make its choice only considering the profit difference.

While DM decision rules are precluded under certain conditions the formation of monopolies, the use of a TC decision rule

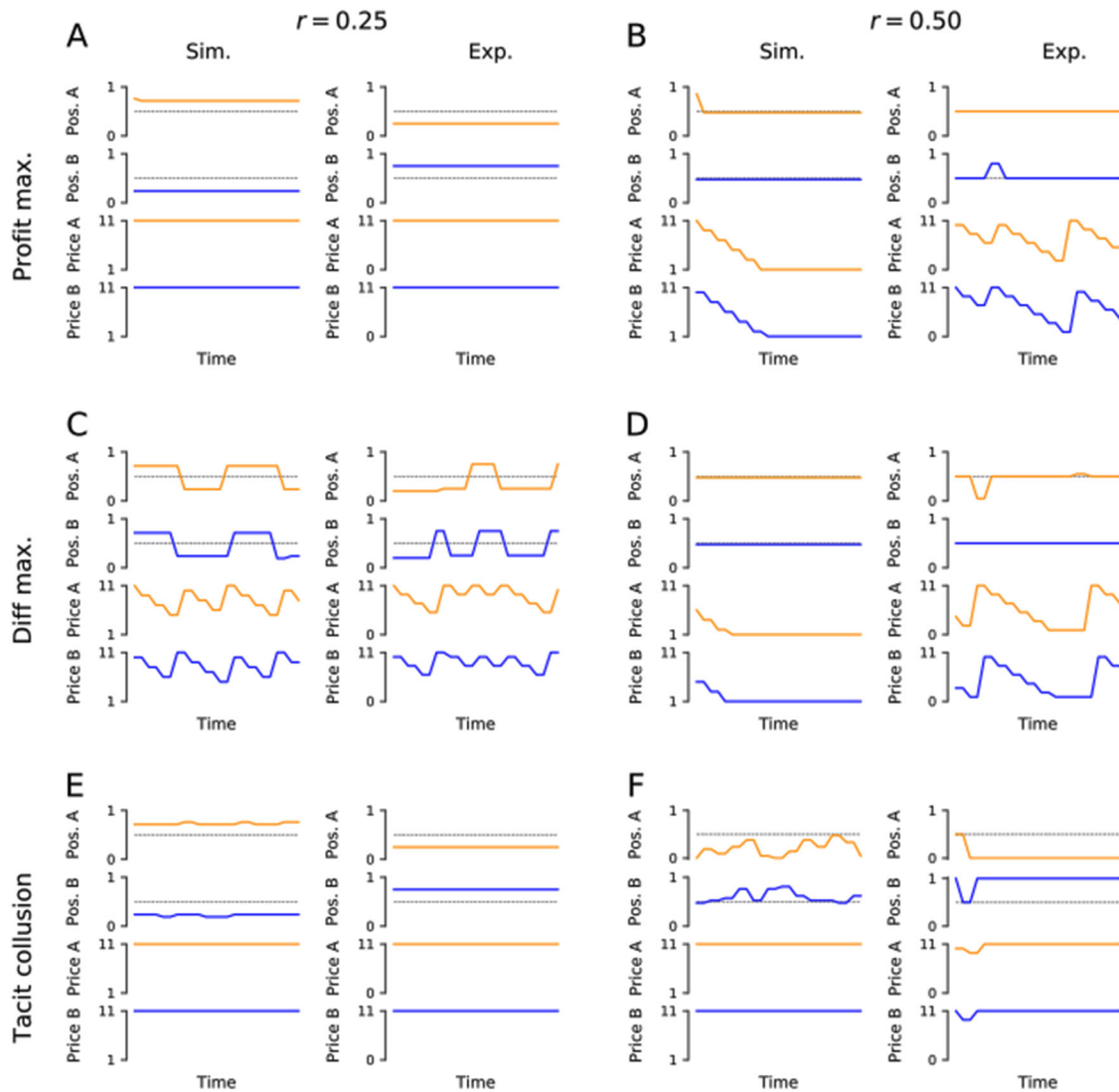


Fig. 4 Analysis of dynamics. Comparison of the dynamics between artificial firms and human controlled firms. Each figure presents the evolution of positions and prices of two firms in competition (orange: Firm A, blue: Firm B; data for artificial firms comes from a single simulation that serves as an example of a typical behavior). **a** Left: artificial firms using a profit maximization strategy; right: two participants with a high score in profit maximization; $r = 0.25$. **b** Same as in subfigure **a** but with $r = 0.50$. **c** Left: two firms using a difference maximization (DM) decision rule; two participants with a high score in DM; $r = 0.25$. **d** Same as in subfigure **c** but with $r = 0.50$. **e** Left: two firms using a tacit collusion (TC) decision rule; right: two participants with a high score in TC; $r = 0.25$. **f** Same as in subfigure **e** but with $r = 0.50$

allowed to relax price competition when information structure was promoting it. As a means to avoid the drawbacks of a competition situation leading to lower profits, the use of such decision rule could be explained by the search for a Pareto optimality (Pareto, 1964) that is to say following the strategies that lead to a distribution of profits such as no firm could earn more, otherwise it would be at the expense of the other. It could also be interpreted as deliberate attempts to emit signals in order to relax competition in a situation where the communication technology needed to lead it rationally is lacking.

Another consideration that is raised by our study is that the consequences of using such decision rules can differ from Nash predictions applied to a basic Hotelling's model under full information: for instance, the use of a TC decision rule under full information leads to maximally differentiate while minimal differentiation would be expected. However, it is now a well-trodden theme that decision rules can be interpreted in terms of their adaptive rationality (Gigerenzer and Reinhard, 2001). As long as a chosen decision rule improves the outcome of the game and

corresponds to relatively stable observable spatial patterns, we can speak of a specific form of rationality arising under the imposed informational constraint. Work by Sutton (1997), applied to the Hotelling's model, explores such an equilibrium notion and weak rationality requirement, based in his case on a single decision rule, which is to seize an opportunity when it presents itself. It would take a further study to understand how the TC decision rule highlighted here indeed constitutes an adaptive rational behavior to informational constraints, either exerted on consumers by means of the availability of information, or exerted on firms by means of their disability to communicate.

For the purpose of our study, we considered a model with a basic architecture. For instance, we used a homogeneous view radius for consumers in our model, mainly for two reasons: (i) it may have been more difficult for the subjects to identify consumers view radius and to adapt their behavior in consequence, (ii) diminish the potential variability of our data. Besides, from a more theoretical point of view, it is equivalent to consider constant radius across the population of consumers as reciprocally

implementing the idea of a limited sphere of influence of the firms. Given this limited amount of vision around consumers or sphere of influence of firms (in the sense, then, of being visible by consumers) it also motivates firms to try to change their location. That being said, It might be relevant to implement heterogeneous view radius in a further study, in order to test the robustness of our results established in a homogeneous setting.

Another implication of our implementation of information is that consumers may be unaware of one of the two options. The unawareness of one of the two options is an abstraction for a consumer that is not willing to inform himself. For instance, we can mention emergency situations (keys lost, car break) where the first solution is picked without consideration of other options. This type of partial attention or restriction to a “consideration set” can also be seen as reflecting a form of incomplete preference relation on the part of the consumers, which has been modeled in different terms in the literature: top options (Rubinstein and Salant, 2011) or consideration sets (Lleras et al., 2017).

Regarding the structure of firm decision-making, we think that a turn-based strategy is more appropriate as it is more likely that a human subject embodying a firm chooses a strategy in reaction to a change of strategy from its competitor. When dealing with rational agents, simultaneous decision-making is possible in the sense that they dispose of full information and unbounded computational abilities, providing them the means to compute the equilibrium and to play accordingly. As we set-up a human subject experiment, we were expecting to deal with non-fully rational agents that are unable to do so. Coordination on price or location policies, leading to a typical situation of maximal/minimal differentiation seems unlikely or at least much more difficult in this configuration. Indeed, pure rationality models such as required to deal with normal forms or extensive forms in experimental game-theory predict behavior to a lesser extent that taking account the incremental feedback of players in repeated sequential situations (Roth and Erev, 1995). We anticipated that subjects will make use of decision rules, leading to more or less stable situations—these decision rules play also the rules of learning heuristics. Then, designing a turn-based game constituted for us a way to facilitate the occurrence of such situations.

From a broader perspective, our results demonstrate interaction effects between consumers and firms’ cognitions, that can deeply impact market dynamics. It, therefore, creates an incentive to think that duopoly regulation should incorporate insights from incomplete markets due to agents limited cognitive abilities. However, most of the focus has been put in behavioral industrial organization to the irrationalities of clients rather than firms. We do not consider our firms irrational either but as constrained to find decision rules in response to their own perception of the consumers’ limited information about themselves. From a theoretical point of view, we are not the first to envision such a problem (Spiegler, 2006). However, the decision rules we stylize and simulate can definitely provide incentives to a more behaviorally oriented approach to duopoly regulation.

Code and data availability

Simulations were implemented using Python and the Python scientific stack (Jones et al., 2001; van der Walt et al., 2011; Hunter, 2007). The code is available at <https://github.com/AurelienNioche/SpatialCompetition>.

The software used for the experimental part of the study is based on a client/server architecture. The client part was developed using the Unity game engine, hosted on a dedicated server and ran in the subjects’ web browser using WebGL API. The code and the assets are available at <https://github.com/AurelienNioche/DuopolyAssets>.

The experiment server was hosted on a dedicated server and developed using the Django framework. The code of the server part is available at <https://github.com/AurelienNioche/DuopolyDjango>.

The analysis program is available at <https://github.com/AurelienNioche/DuopolyAnalysis>. Figures 3 and 4 were produced using raw data that are available at <https://github.com/AurelienNioche/DuopolyAnalysis>.

Received: 26 September 2018 Accepted: 28 February 2019

Published online: 26 March 2019

References

- Amir O, David GR, Yaakov KG (2012) Economic games on the internet: The effect of \$1 stakes. *PLoS ONE* 7(2):e31461. <https://doi.org/10.1371/journal.pone.0031461>
- Barreda-Tarrazona I, Aurora G-G, Nikolaos G, Joaquin A-F, Agustin G-S (2011) An experiment on spatial competition with endogenous pricing. *Int J Ind Organ* 29(1):74–83. <https://doi.org/10.1016/j.ijindorg.2010.02.001>
- Brenner S (2005) Hotelling games with three, four, and more players. *J Reg Sci* 45(4):851–64. <https://doi.org/10.1111/j.0022-4146.2005.00395.x>
- Brown S (1989) Harold hotelling and the principle of minimum differentiation. *Progress Human Geogr* 13(4):471–93. <https://doi.org/10.1177/030913258901300401>
- Buhrmester M, Tracy K, Samuel DG (2011) Amazons mechanical turk. *Perspect Psychol Sci* 6(1):3–5. <https://doi.org/10.1177/1745691610393980>
- Cremer H, Maurice M, Jacques-François T (1991) Mixed oligopoly with differentiated products. *Int J Ind Organ* 9(1):43–53. [https://doi.org/10.1016/0167-7187\(91\)90004-5](https://doi.org/10.1016/0167-7187(91)90004-5)
- d’Aspremont C, Jaskold Gabszewicz J, Thisse J-F (1979) On hotellings “stability in competition”. *Econometrica* 47(5):1145. <https://doi.org/10.2307/1911955>
- Dudey M (1990) Competition by choice: The effect of consumer search on firm location decisions. *Am Econ Rev* 80(5):1092–1104. <https://ideas.repec.org/a/aea/aecrev/v80y1990i5p1092-1104.html>
- Eaton BC, Richard GL (1975) The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition. *Rev Econ Stud* 42(1):27. <https://doi.org/10.2307/2296817>
- Economides N (1993) Hotelling’s “main street” with more than two competitors. *J Reg Sci* 33(3):303–19. <https://doi.org/10.1111/j.1467-9787.1993.tb00228.x>
- Gigerenzer G, Reinhard S (eds.) (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press. <https://mitpress.mit.edu/books/bounded-rationality>
- Hotelling H (1929) Stability in competition. *Econ J* 39(153):41. <https://doi.org/10.2307/2224214>
- Huck S, Wieland M, Nicolaas JV (2002) The east end, the west end, and king’s cross: On clustering in the four-player hotelling game. *Econ Inq* 40(2):231–40. <https://doi.org/10.1093/ei/40.2.231>
- Hunter JD (2007) Matplotlib: A 2D graphics environment. *Comput Sci Eng* 9(3):90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Ipeirotis PG (2010) Demographics of mechanical turk. Ce-DER-10-01. New York University, New York, NY, <http://www.ipeirotis.com/wp-content/uploads/2012/02/CeDER-10-01.pdf>
- Irmen A, Jacques-François T (1998) Competition in multi-characteristics spaces: Hotelling was almost right. *J Econ Theory* 78(1):76–102. <https://doi.org/10.1006/jeth.1997.2348>
- Jones E, Travis O, Pearu P (2001) SciPy: Open source scientific tools for python. <http://www.scipy.org>
- Kahneman D (2003) A perspective on judgment and choice: Mapping bounded rationality. *Am Psychol* 58(9):697–720. <https://doi.org/10.1037/0003-066x.58.9.697>
- Kephart C, Friedman D (2015) Hotelling revisits the lab: Equilibration in continuous and discrete time. *J Econ Sci Assoc* 1(2):132–45. <https://doi.org/10.1007/s40881-015-0009-z>
- Kruse JB, David JS (2000) Location, cooperation and communication: an experimental examination. *Int J Ind Organ* 18(1):59–80. [https://doi.org/10.1016/S0167-7187\(99\)00034-x](https://doi.org/10.1016/S0167-7187(99)00034-x)
- Lleras JS, Masatlioglu Y, Nakajima D, Ozbay EY (2017) When more is less: Limited consideration. *J Econ Theory* 170:70–85. <https://doi.org/10.1016/j.jet.2017.04.004>
- Loertscher S, Muehlheusser G (2011) Sequential location games. *RAND J Econ* 42(4):639–63. <https://doi.org/10.1111/j.1756-2171.2011.00148.x>
- Meegan DV (2010) Zero-sum bias: perceived competition despite unlimited resources. *Front Psychol* 1:191. <https://doi.org/10.3389/fpsyg.2010.00191>
- Paolacci G, Jesse C, Panagiotis GI (2010) Running experiments on amazon mechanical turk. *Judgm Decis Mak* 5(5):411–19. <https://EconPapers.repec.org/RePEc:jdj:journl:v:5:y:2010:i:5:p:411-419>

- Pareto V (1964). *Cours d'économie Politique*. vol 1. Librairie Droz. <https://www.cairn.info/cours-d-economie-politique-tomes-1-et-2--9782600040143.htm>
- Prescott EC, Visscher M (1977) Sequential location among firms with foresight. *Bell J Econ* 8(2):378. <https://doi.org/10.2307/3003293>
- Rózycka-Tran J, Paweł B, Bogdan W (2015) Belief in a zero-sum game as a social axiom: A 37-nation study. *J Cross-Cult Psychol* 46(4):525–48. <https://doi.org/10.1177/0022022115572226>
- Roth AE, Erev I (1995) Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games Econ Behav* 8(1):164–212. [https://doi.org/10.1016/S0899-8256\(05\)80020-X](https://doi.org/10.1016/S0899-8256(05)80020-X)
- Rothschild A (1947) Price theory and oligopoly. *Econ J* 57(227):299–320. <https://doi.org/10.2307/2225674>
- Rubinstein A (1991) Comments on the interpretation of game theory. *Econometrica* 59(4):909–924. <https://doi.org/10.2307/2938166>
- Rubinstein A, Salant Y (2011) Eliciting welfare preferences from behavioural data sets. *Rev Econ Stud* 79(1):375–387. <https://doi.org/10.1093/restud/rdr024>
- Schultz C (2009) Transparency and product variety. *Econ Lett* 102(3):165–68. <https://doi.org/10.1016/j.econlet.2008.12.008>
- Spiegler R (2006) Competition over agents with boundedly rational expectations. *Theor Econ* 1(2):207–231. <https://econtheory.org/ojs/index.php/te/article/viewArticle/20060207>
- Stahl K (1982) Differentiated products, consumer search, and locational oligopoly. *J Ind Econ* 31(1/2):97. <https://doi.org/10.2307/2098007>
- Stigler GJ (1961) The economics of information. *J Political Econ* 69(3):213–25. <https://doi.org/10.1086/258464>
- Sutton J (1997) One smart agent. *RAND J Econ* 28(4):605. <https://doi.org/10.2307/2555778>
- Thaler R (1980) Toward a positive theory of consumer choice. *J Econ Behav Organ* 1(1):39–60. [https://doi.org/10.1016/0167-2681\(80\)90051-7](https://doi.org/10.1016/0167-2681(80)90051-7)
- Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–31. <https://doi.org/10.1126/science.185.4157.1124>
- van der Walt S, Colbert SC, Gaël V (2011) The NumPy array: A structure for efficient numerical computation. *Comput Sci Eng* 13(2):22–30. <https://doi.org/10.1109/mcse.2011.37>
- Webber MJ (1972) *The impact of uncertainty upon location*. MIT Press, Cambridge, MA. <https://mitpress.mit.edu/books/impact-uncertainty-location>

Acknowledgements

This work was supported by the Agence Nationale de la Recherche (ANR-16-CE38-0003). The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

Author contributions

AN and BG wrote the code, performed the experiments and the data analysis; AN, BG, TB, NR, and SB-G designed the study and co-wrote the manuscript.

Additional information

Supplementary information: The online version of this article (<https://doi.org/10.1057/s41599-019-0241-x>) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://www.nature.com/reprints>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

RÉSUMÉ

La valeur subjective est une construction théorique omniprésente dans l'étude de la prise de décision. Dans cette littérature, les décisions des individus sont souvent conçues selon un processus en deux étapes. Ils attribuent d'abord des valeurs aux options disponibles, puis choisissent l'option ayant la valeur la plus élevée.

Aussi, deux manières de construire des valeurs subjectives sont souvent envisagées : par description et par expérience. Apprendre par description correspond à obtenir des informations sur la valeur des options disponibles, via des représentations visuelles et sémantiques. Par exemple, lorsque que l'on sélectionne un restaurant sur internet, on peut se référer aux notes et commentaires fournis par les autres utilisateurs. A l'inverse, apprendre par expérience correspond à construire des valeurs subjectives par essais-erreurs. Cela correspondrait donc à essayer différents restaurants, et ainsi se construire une appréciation de ces derniers.

Dans cette thèse, nous cherchons à évaluer si les valeurs construites via ces deux méthodes sont commensurables. Plus précisément, nous cherchons à établir si les individus sont capables de comparer les valeurs acquises par expérience et description, et si oui, par quels processus de décision.

MOTS CLÉS

Prise de décision, Modélisation, Neurosciences Computationnelles, Apprentissage par renforcement, Incertitude, Description et Expérience

ABSTRACT

Subjective value is an ubiquitous theoretical construct in the study of decision making. In this literature, individuals' decisions are often conceived in a two-step process. They first assign values to the available options, and then choose the option with the highest value.

Thus, two ways of constructing subjective values are often considered: by description and by experience. Learning by description corresponds to obtaining information about the value of available options, via visual and semantic representations. For example, when selecting a restaurant on the Internet, we can refer to the ratings and comments provided by other users. Conversely, learning by experience corresponds to building subjective values by trial and error. This would correspond to trying different restaurants, and thus building an appreciation of them.

In this thesis, we seek to evaluate whether the values constructed via these two methods are commensurable. More precisely, we seek to establish whether individuals are able to compare values learned by experience and description, and if so, by which decision processes.

KEYWORDS

Decision-making, Modeling, Computational neuroscience, Reinforcement learning, Uncertainty, Description-experience gap