DISS. ETH NO 24698

NEW RISK MANAGEMENT FRAMEWORKS:

FROM QUANTUM DECISION THEORY TO SYSTEM RESILIENCE

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

TATYANA KOVALENKO

M.Sc. in Economics, Belarus State Economic University (BSEU)

born on 25.06.1986

Citizen of the Republic of Belarus

accepted on the recommendation of

Prof. Dr. Didier Sornette, examiner

Prof. Dr. Hans Rudolf Heinimann, co-examiner

2017

Abstract

This dissertation is a compilation of publications and publication manuscripts that seek to improve existing risk management approaches from two perspectives: (i) at a macro-level, by developing a general framework for risk and resilience management, and (ii) at a micro-level, by modeling individual and aggregate risky choices within a novel probabilistic Quantum decision theory (QDT).

The first publication proposes an operational definition of resilience, seeing it as a measure of stress that is complementary to the risk measures. Distinguishing between stressors (exogenous and endogenous forces acting on the system) and stress (reaction of the system), we discuss systems' dynamics under different environmental and stress conditions. We suggest a four-level resilience hierarchy. With focus on socio-economic systems, strategic principles for resilience build-up, (human) limitations and original operational solutions are delineated.

The second publication introduces four risk and resilience management regimes, which are identified based on (i) the level of stress induced by environmental exogenous demands or endogenous processes and (ii) the degree of uncertainty/predictability of a system. We refer to this framework as "4 quadrants" of risk severity and system control. Corresponding response mechanisms and management instruments are outlined.

In the third publication manuscript, we investigate a probabilistic approach to modeling individual and aggregate binary risky choices, and present the first calibration of QDT to empirical dataset. We demonstrate that a simple probabilistic model, without adjustable parameters, can account for the majority of choice reversals between two repetitions of the experiment, and can be further refined by introducing heterogeneity: differentiation of decision makers into "overconfident" and "contrarian". This supports the fundamental tenet of QDT, which models choice as an inherent probabilistic process, such that the probability of a prospect is expressed as the sum of its utility and attraction factors. We parameterize (a) the utility f-factor with a stochastic cumulative prospect theory (logit-CPT), and (b) the attraction q-factor with a constant absolute risk aversion function, which captures aversion to large losses. The QDT model outperforms the logit-CPT. Our quantitative analysis supports the existence of an intrinsic limit of predictability associated with the inherent probabilistic nature of choice.

Finally, the fourth publication manuscript initiates a data-driven exploration of the underlying theoretical construct of QDT. A novel QDT interpretation of the conjunction fallacy exposes the state of mind of a decision maker as a distinct source of uncertainty and interference effects. We link typicality judgements to probability amplitudes of the decision modes in the state of mind, and quantify the level of uncertainty and the relative contributions of prospect's interfering modes to the resultant probability judgement. This enables inferences about the QDT attraction (interference) q-factor for different prospects (compatible/incompatible) and varying uncertainty levels. Under high uncertainty, the q-factor tends to converge to the negative range $q \in (-0.25, -0.15)$. This hypothesized universal "aversion" q is independent of the (un-)attractiveness of a prospect under more certain conditions, which distinguishes it from the previously considered QDT "quarter law". The universal "aversion" q substantiates the heuristic QDT "uncertainty aversion principle" and provides a theoretical basis for modeling different risk attitudes, such as aversions to uncertainty, to risk or to losses. Empirically motivated, we consider a novel "QDT indeterminacy principle", as a fundamental limit of the precision with which certain sets of prospects can be simultaneously assessed or elicited.

Résumé

Cette dissertation est une compilation de publications parues et de manuscripts pour publications qui cherchent à améliorer les approches existantes de la gestion des risques sous deux angles: (i) à un niveau macro, par l'élaboration d'un cadre général pour la gestion des risques et de la résilience, et (ii) à un niveau micro, par la modélisation de choix risqués individuels et globaux à l'aide d'une nouvelle théorie de la décision quantique probabiliste (QDT).

La première publication propose une définition opérationnelle de la résilience, en la considérant comme une mesure de stress complémentaire aux mesures de risque. Distinguer entre les facteurs de stress (forces exogènes et endogènes agissant sur le système) et le stress (réaction du système), nous discutons la dynamique des systèmes dans différentes conditions environnementales et de stress. Nous proposons une hiérarchie de la résilience possédant quatre niveaux. En mettant l'accent sur les systèmes socio-économiques, on délimite les principes stratégiques pour l'établissement de la résilience, les limitations (humaines) existantes et des solutions opérationnelles originales.

La deuxième publication présente les quatre régimes de gestion des risques et de la résilience, qui sont identifiés en fonction (i) du niveau de stress induit par les demandes exogènes environnementales ou les processus endogènes et (ii) du degré d'incertitude et de prévisibilité d'un système. On réfère à ce cadre comme étant celui des "4 quadrants" de la gravité des risques et du degré de contrôle possible du système. Les mécanismes sous-jacents des réponses possibles et les instruments de gestion sont aussi décrits.

Dans le troisième manuscript, nous étudions une approche probabiliste de la modélisation des choix binaires au niveau de chaque individu et au niveau agrégé et nous présentons la première calibration de la QDT à un ensemble de données empiriques. Nous démontrons qu'un modèle probabiliste simple, sans paramètre ajustable, peut décrire la majorité des inversions de choix entre deux répétitions de l'expérience. Ce modèle peut être affiné par l'introduction d'une différenciation entre des décideurs "trop confiants" et des décideurs "contrariants". Ce résultat supporte le principe fondamental de la QDT, qui modélise les choix comme étant probabilistes de manière inhérente, de sorte que la probabilité d'un prospect est exprimée comme la somme de ses facteurs d'utilité et d'attraction. Nous paramétrons (a) le facteur f de l'utilité avec une version stochastique de la théorie des prospects cumulatifs (logit-CPT), et (b) le facteur d'attraction q avec avec une function d'aversion relative constante au risque qui représente l'aversion à de grandes pertes. On trouve que le modèle QDT est supérieur au model logit-CPT. Notre analyse quantitative soutient l'existence d'une limite intrinsèque à la prevision, limite qui résulte de la nature probabiliste inhérente des choix.

Enfin, le quatrième manuscrit présente une exploration des bases fondamentales de la construction théorique de la QTD. Une nouvelle interprétation basée sur la QDT du paradoxe du biais de représentativité met l'accent sur l'importance de l'état d'esprit d'un décideur comme une source distincte d'incertitude et d'interférences. Nous associons les jugements d'une caractéristique typique aux amplitudes de probabilité des modes de décision dans l'état d'esprit d'un décideur. Nous quantifions le niveau d'incertitude et les contributions relatives aux modes interférants des prospects au jugement de la probabilité résultante d'un choix donné. Cela permet de déduire des informations précieuses concernant le facteur d'attraction q de la QDT pour différents types de prospects, qu'ils soient du type compatible ou incompatible et en fonction de different niveaux d'incertitude. En présence d'une forte incertitude, le facteur q tend à converger dans l'intervalle négatif (-0.25; -0.15). Cette "aversion" universelle que nous conjecturons est indépendante de l'attrait ou répulsion d'un prospect sous conditions de plus grande certitude, ce qui la distingue de la "loi du quart" de la QDT qui avait été précédemment proposée. L'aversion universelle q justifie l'hypothèse d'un "principe d'aversion à l'incertitude" et fournit une base théorique pour la modélisation de différentes attitudes au risque, telles que les aversions à l'incertitude, aux risques ou aux pertes. Empiriquement motivé, nous introduisons un nouveau "principe d'indétermination" de la QDT, qui est présenté comme une limite fondamentale de la précision avec laquelle certains ensembles de prospects peuvent être évalués ou obtenus simultanément.

Acknowledgments

ETH Zurich and the Chair of Entrepreneurial Risks are a community of enthusiastic, dedicated and talented people. I am fortunate and grateful for the opportunity to join them in the scientific exploration of the world and myself.

First and foremost, I am grateful to Prof. Didier Sornette for this incredible life adventure. He is an exceptional mentor and leader, generously sharing his excellent expertise, original ideas and time. He is a true *Homo Ludens*, whose drive, curiosity and optimism spark off, encourage and transform hard work into a play, at least for a while. I thank you for knowledge, patience and support.

I would like to thank Prof. Vyacheslav Yukalov and Dr. Elizaveta Yukalova. Their fruitful collaborative work with Prof. Sornette laid the foundation of the Quantum decision theory, which I rely on in decision making research and practice. Vyacheslav and Elizaveta are an inspiring example of similar-minded, complementary and supportive partners.

I deeply appreciate Prof. Hans Rudolf Heinimann being co-examiner of my work, and Prof. Antoine Bommier presiding over the examination committee. Their professional expertise matches perfectly the two research areas - resilience and risky decision making, - and their feedback is highly valued.

I am very grateful to Dr. Ryan O. Murphy for making an empirical dataset of risky decisions available for our analysis. I know him as a gifted teacher and a thorough researcher, whose comprehensive advice and cheerful support I appreciate.

Special words of appreciation are dedicated to Sabine Vincent, Morgan Siffert and Emmanuel Munich - master students at the Chair of Entrepreneurial Risks, whom I was happy to work with. I thank you for the tenaciousness, imagination and fostering my supervising skills.

I wish to thank Dr. Igor Linkov for inviting me to the research workshop on ”Resilience-based approaches to critical infrastructure safeguarding” (Azores, Portugal, 2016). A fine organization facilitated collective brainstorming between academics and practitioners. I thank Dr. Ivo Häring, Prof. José Manuel Palma-Oliveira, Marie-Valentine Florin, Prof. Hans Rudolf Heinimann, Prof. Giovanni Sansavini, Dr. Maksim Kitsak and other participants of the workshop for productive discussions on resilience.

Back to ETH Zurich, I thank Dr. Maroussia Favre and Dr. Robert ten Brincke, my fellow researchers in decision theory, for useful exchanges. Current and former members of the Chair, many have already deserved the Dr. title, contributed to my everyday working experience, among them Afina Inina, Diego Ardila, Richard Senner, Vladimir Filimonov, Guilherme Demos, Spencer Wheatley, Michael Schatz, Dorsa Sanadgol, Zalán Forró, Hyun-U Sohn, Maciej Jagielski, Susanne von der Becke, Lucas Fivet, Georges Harras, Dmitry Chernov, Ryohei Hisano, Ke Wu. I am grateful to Yannick Malevergne, Peter Cauwels, Qunzhi Zhang and David Solo for their senior advice and inspiration. Isabella Bieri, Heidi Demuth and Judith Holzheimer ensured a smooth, efficient and cheerful administrative support.

I thank anonymous reviewers and editors for the assistance in improving manuscripts.

This work was partially supported by the Swiss National Foundation, under grant 105218 159461 for the project on “Quantum Decision Theory”.

Dedicated to my loving family and friends: Natallia, Mihail, Ivan, Julia, Nicole, Christian, Olga and Elizaveta.

Zurich, October 2017 *Tatyana Kovalenko*

To my parents, Natallia and Mihail.

Contents

1 Introduction and motivation **1**
1.1 Risk and resilience 2
1.2 Quantum decision theory 3

2 Risk and resilience management **5**
Kovalenko, T. and Sornette, D. (2013): Dynamical diagnosis and solutions for resilient natural and social systems. In: *Planet@Risk*, 1 (1): 7-33, Davos: Global Risk Forum GRF Davos. 6
Kovalenko, T. and Sornette, D. (2016): Risk and Resilience Management in Social-Economic Systems. In Florin, M.-V. and Linkov, I. (Eds.): *IRGC Resource Guide on Resilience*, Lausanne: EPFL International Risk Governance Center. 34
Häring, I. and Sansavini, G. and Bellini, E. and Martyn, N. and Kovalenko, T. and Kitsak, M. and Vogelbacher, G. and Ross, K. and Bergerhausen, U. and Barker, K. and Linkov, I. (2017): Towards a generic resilience management, quantification and development process: General definitions, requirements, methods, techniques and measures, and Case studies. In Linkov, I. and Palma-Oliveira, J.M. (Eds.): *Resilience and Risk: Methods and Application in Environment, Cyber and Social Domains*, Dordrecht: Springer. 45

3 Quantum decision theory parametrization **49**
Vincent, S. and Kovalenko, T. and Yukalov, V.I. and Sornette, D. (2017): Calibration of Quantum decision theory, aversion to large losses and predictability of probabilistic choices. In: *Submitted to Theory and Decision*, Springer. 50
Kovalenko, T. and Sornette, D. (2017): Conjunction fallacy in Quantum decision theory. In: *Working paper* 103

4 Conclusion **149**

1 Introduction and motivation

Life in changing environment is inseparably linked to risk. Therefore the aspiration to survive and prosper stipulates the necessity to understand and forecast potential threats and their consequences. Standard approaches to risk often relied on static statistical measures, such as value at risk, providing a "snapshot" view of a system. The danger is, however, that this rigid approach to risk may be blind to numerous changing conditions, such as slow-moving risks and maturing instability of a system. Even worse, inadequate metrics can generate false perception of safety and misdiagnose unsustainable trends.

A dynamical approach to risk management emphasizes the continuous quantification and monitoring of risk factors, their interconnections and influence on a functioning system. The development of a dynamical paradigm naturally turns one's attention towards the system itself - its ability to respond to new environmental demands, its capacity to withstand disturbances and disruptions, to adapt and transform. This puts resilience under the limelight.

Why some systems are more resilient than others? How can relevant resilience properties be designed and enhanced? Pushing to the limit, can a system be risk-proof, i.e. invulnerable in ambiguous, unpredictable environment, and benefit from any type of variability? Or, more realistically, can some of the risks be transformed into opportunities? This questions motivate our (re)search of resilience.

Resilience implies *re*action to a risk factor, thus it is the feature of a "living" active system. The property of resilience can either be governed towards and managed at a macro-level, or emerge from interactions of individual agents at a micro-level. In both cases, resilience is tightly connected with the social component of a system and decision making process.

Since the mid-twentieth century, theoreticians and experimentalists from economics and psychology made significant efforts to document discrepancies between normative and observed choice behavior. In this way, conventional decision theory contraposes prescriptive models, which are based on expected value or expected utility (EU-type), to descriptive models (nonEU-type). The latter is a medley of behavioral approaches that are usually conceived as an explanation of a particular identified bias, or several of them.

Variability of choice is a well-known and ubiquitously observed pattern. It is reported under different conditions, both as a heterogeneity within a group, as well as variations of an individual response within a repeated setting. Surprisingly, this characteristic feature of choice behavior is often ignored, disregarded, treated as erroneous and mistaken.

Our decision making research is motivated by the question: is choice intrinsically probabilistic? And if so, what factors do affect choice probabilities? Quantum decision theory is instrumental in this quest. It is a probabilistic choice theory that naturally incorporates the influence of interfering factors. Moreover, it can be reduced to conventional decision theories, thus allowing for straightforward model comparison.

This thesis comprises four self-contained research articles with relevant literature reviews. They fall into two parts - risk and resilience management, and quantum decision theory. The rest of the Introduction section outlines the objectives and gives an overview of the conducted research.

1.1 Risk and resilience

The main objective of this research strand is to develop a systemic view on risk and resilience - a general management framework that is relevant for an arbitrary system, with an emphasis on socio-economic and financial systems.

Since the 1990s, a more systematic quantitative approach to risk management was developed for practical implementation in finance and in many scientific and industrial areas. Lately, the concept of resilience spread its influence from engineering, social (e.g. psychology) and natural (e.g. ecology) sciences to management, economics and finance. This new broader application of resilience calls for a reexamination of the previously developed methodology, its adaptation to the new fields of interest and the design of technics to foster resilience of social-economic systems.

A generic resilience approach is in an active stage of formation, where multidisciplinary elements of methodology and practice are being tried on, fused and re-fused, expelled or merged within the core framework (which we determine as 'system' 'dynamics'). Unsurprisingly, researchers and practitioners tend to view resilience being refracted through the lens of their discipline or regarding a specific system in consideration. So, an engineer would emphasize resistance property, safety and robustness of a structure; an ecologist - capacity of a system to respond to a perturbation or disturbance, its sustainability; a manager - business continuity, etc. Inclusive relations between involved methodological concepts vary and are often inconsistent. For example, a risk specialist could classify resilience as one of the risk management strategies that is especially relevant in a highly uncertain and ambiguous environment. In contrast, a resilience specialist would consider risk management processes (risk identification, assessment and control) as a part of the extensive resilience management.

A reconcilement of the resilience and risk management approach is, as a "red thread", traced through the first publication (**Kovalenko, T.** and Sornette, 2013) **(1)**. Recognition of the central role of the "stress" concept allows positioning risk and resilience as its complementary measures. Further investigation is required to determine whether a system can benefit from a stressor, and, at the limit, from all possible stressors, i.e., is there "antifragility" beyond resilience? Based on the literature review and case study, we propose a four-level resilience hierarchy and draw generic recipes for building up resilience.

A synthesis of the proposed view on risk and resilience, connected by the concept of "stress", gave rise to a novel management framework. We refer to it as the "4 quadrants" of risk severity and system control (**Kovalenko, T.** and Sornette, 2016) **(2)**. Response mechanisms of a system in each regime are outlined, as well as relevant management instruments.

This part is concluded with a discussion on the correspondence between the two key propositions: a four-level resilience hierarchy and the "4 quadrants" of risk severity and system control. The former enriches our risk and resilience management and completes the unified risk-resilience (R-R) approach. The practical application and deployment of a holistic R-R management system may be facilitated by standardization or resilience management processes, on par with risk management, creation of a taxonomy of methods and detailed case studies (Häring et al., 2017) **(3)**.

Among important aspects of resilience generation, we should mention: (i) establishing clear goals and right incentives, (ii) promoting heterogeneity and individual strength, (iii) overcoming intrinsic human limits and biases and (iv) facilitating collective action and collaboration. These topics provide additional motivation for the subsequent research line on decision theory.

Publications

(1) **Kovalenko, T.** and D. Sornette. Dynamical diagnosis and solutions for resilient natural and social systems. *Planet@Risk*, 1(1):7–33, 2013

(2) **Kovalenko, T.** and D. Sornette. Risk and resilience management in social-economic systems. In I. Linkov and M.-V. Florin, editors, *IRGC Resource Guide on Resilience*. EPFL International Risk Governance Center, Lausanne, 2016

(3) I. Häring, G. Sansavini, E. Bellini, N. Martyn, **Kovalenko, T.**, M. Kitsak, G. Vogelbacher, K. Ross, K. Bergerhausen, U. and Barker, and I. Linkov. Towards a generic resilience management, quantification and development process: General definitions, requirements, methods, techniques and measures, and case studies. In I. Linkov and J.M. Palma-Oliveira, editors, *Resilience and Risk: Methods and Application in Environment, Cyber and Social Domains*. Springer, Dordrecht, 2017

1.2 Quantum decision theory

Quantum decision theory (QDT) interprets decisions as intrinsically probabilistic. This means that observed variations in choices are not treated as errors, anomalies or exceptions, but rather considered to reveal a true stochastic nature of choice. QDT utilizes the mathematics of Hilbert spaces and some of the formalism originated from quantum mechanics. It allows one to account for uncertainty and to explain paradoxes of "classical" decision theories via quantum-like effects in decision processes. Such effects include interferences between choice alternatives (prospects) and the entanglement of a decision-maker's state of mind.

As a probabilistic framework, QDT assigns to each alternative (a prospect π_j) in a decision making problem a certain probability $p(\pi_j)$ of the prospect to be chosen. Technically, this probability is defined as the average value of a prospect's operator with respect to a decision-maker's state of mind, which is also represented as an operator. Quantification of these operators (for humans) is extremely challenging. It consists in (noninvasive) elicitation of weights (i.e. squared probability amplitudes) of context-dependent decision modes. The task is even more complicated due to the time-dependence of both operators. These difficulties explain why, until now, this underlying theoretical construct was not applied directly to model choice behavior.

Fortunately, there is an indirect way. It is based on the most general QDT relation that represents prospect's probability $p(\pi_j)$ as a sum of a two factors - its utility $f(\pi_j)$ and attraction $q(\pi_j)$:

$$p(\pi_j) = f(\pi_j) + q(\pi_j) \; .$$

The following constraints are applied:

- the probability $p(\pi_j) > 0$ and normalized across all N alternatives $\sum_{j=1}^{N} p(\pi_j) = 1$;
- the utility factor follows classical probability rules, thus $f(\pi_j) > 0$ and $\sum_{j=1}^{N} f(\pi_j) = 1$;
- the attraction factor $q(\pi_i) \in [-1; 1]$ and follows an alternation rule with $\sum_{j=1}^{N} q(\pi_j) = 0$.

The attraction q-factor is the principal novel ingredient of QDT, which captures interference effects. Theoretically, the functional form of both $f(\pi_i)$ and $q(\pi_i)$ is very flexible. It can include different conventional decision models (EU- or nonEU-type) or alternative formulations, as a function of the parameters defining (sets of) prospects and dependent on context and framing. Despite its simplicity, the indirect way is useful and provides new testable quantitative predictions. The main prediction is called the "QDT quarter law". It suggests, with no prior assumptions, an average value of $q(\pi_i) = \pm 1/4$ for a binary choice (between two prospects).

It is important to stress that previous data analysis within QDT was confined to the formulation of the utility f-factor as the ratio of prospects' expected values, and the calculation of attraction the q-factor as a difference between observed choice probabilities (frequencies) and the above mentioned f-factor. This approach does not involve parameters, thus assumes an homogeneous population. It also attributes all subjective risk attitudes and other possible influencing factors (both persistent and momentary) to one attraction q-factor. On a positive side, this analysis is simple, robust and on many occasions demonstrated the general agreement of data with the prediction of the "QDT quarter law".

The main objectives for this part of my research are:

- to reexamine the evidence fro a *probabilistic* nature of decision making;
- to *parameterize* QDT based on the general representation of a prospect probability as a sum of the f and q factors (an indirect way);
- to attempt an in-depth empirical analysis that involves the *underlying* QDT mechanism in order to trace quantum-like effects, interference and entanglement, in action (a direct way).

The first article on decision making (Vincent et al., 2017) **(4)** analyses a mid-size experimental dataset of binary risky choices. Data analysis supports the probabilistic approach to modeling choice behavior, and indicates the existence of intrinsic limits of its predictability. We suggest that stochastic decision making can provide evolutionary advantage, for coping with adverse external and internal factors in complex environment. We propose a QDT parametrization based on a stochastic version of cumulative prospect theory (for the utility f-factor) and a constant absolute risk aversion function (for the attraction q-factor). This corresponds to separating aversion to large losses as an interfering effect. We successfully calibrate this QDT model on both an ensemble of individuals and single decision makers.

The final article (**Kovalenko, T.** and Sornette, 2017) **(5)** turns back to pure QDT fundamentals. We endeavor to understand decision making processes in details, and evoke an exemplary conjunction fallacy for that purpose. We decompose the entanglement process in the state of mind stepwise. It highlights the effect of framing during pre-exposition of a decision maker to the description of a subject, e.g. the famous 'Linda'. This phase is at the origin of uncertainty and interference effects. Assuming several extreme parametric formulations, we are able to analyze the relative influence of interfering decision modes on the prospect probability (probability judgement). This data-driven approach has led us to a new fundamental perspective: an universal "aversion" q, and possible limits of simultaneous inferences with respect to certain types of prospects.

Publication manuscripts

(4) S. Vincent, **Kovalenko, T.**, V.I. Yukalov, and D. Sornette. Calibration of quantum decision theory, aversion to large losses and predictability of probabilistic choices. *Submitted to Theory and Decision*, 2017

(5) **Kovalenko, T.** and D. Sornette. Conjunction fallacy in quantum decision theory. *Working paper*, 2017

2 Risk and resilience management

This chapter presents publications as originally published, reprinted with permission from the corresponding publishers. The copyright of the original publications is held by the respective copyright holders, see the following copyright notices.

The discussion section of this chapter contains illustrations from a book, reproduced with permission from the corresponding publisher. The copyright of the original publications is held by the respective copyright holder, see the following copyright notice.

GRF Davos Planet@Risk, Volume 1, Number 1, October 2013

http://www.planet-risk.org

Dynamical Diagnosis and Solutions for Resilient Natural and Social Systems

KOVALENKO, Tatyana[a], SORNETTE, Didier[b]

[a] Department of Management, Technology and Economics, ETH Zurich, Switzerland, email: tkovalenko@ethz.ch
[b] Department of Management, Technology and Economics, ETH Zurich, Switzerland, email: dsornette@ethz.ch

Abstract – The concept of resilience embodies the quest towards the ability to sustain shocks, to suffer from these shocks as little as possible, for the shortest time possible, and to recover with the full functionalities that existed before the perturbation. We propose an operational definition of resilience, seeing it as a measure of stress that is complementary to the risk measures. Emphasis is put on the distinction between stressors (the forces acting on the system) and stress (the internal reaction of the system to the stressors). This allows us to elaborate a classification of stress measures and of the possible responses to stressors. We emphasize the need for characterizing the goals of a given system, from which the process of resilience build-up can be defined. Distinguishing between exogenous versus endogenous sources of stress allows one to define the corresponding appropriate responses. The main ingredients towards resilience include (1) the need for continuous multi-variable measurement and diagnosis of endogenous instabilities, (2) diversification and heterogeneity, (3) decoupling, (4) incentives and motivations, and (5) last but not least the (obvious) role of individual strengths. Propositions for individual training towards resilience are articulated. The concept of "crisis flight simulators" is introduced to address the intrinsic human cognitive biases underlying the logic of failures and the illusion of control, based on the premise that it is only by "living" through scenarios and experiencing them that decision makers make progress. We also introduce the "time@risk" framework, whose goal is to provide continuous predictive updates on possible scenarios and their probabilistic weights, so that a culture of preparedness and adaptation be promoted. These concepts are presented towards building up personal resilience, resilient societies and resilient financial systems.

Keywords – *resilience, stress, stressor, failure, human cooperation, antifragility, illusion of control, crisis flight simulator, time@risk*

1. Introduction

Interesting systems are out-of-equilibrium and subjected to external influences. In biology, the only true equilibrium state is death (Selye, 1973). In contrast, living organisms are remarkable engines that use energy and matter to generate internal order and external entropy. Being coupled to some outside environment, any interesting biological or social system is under the influence of fluxes, their fluctuations and trends as well as perturbations of various types (Lipsitz, 2002). Under these exogenous influences, they organize endogenously, attempting to self-propagate, grow and invade all available niches. These systems attempt to stabilize, at least for a time, towards some sort of dynamical equilibrium or are managed to stay close to a desirable state. Nevertheless, numerous exogenous and endogenous stress-factors continuously destabilize these systems. An outstanding question, which is increasingly crucial to modern human societies, is how to ensure survivability, sustainability, resilience as well as promise of better well-being and happiness in the presence of the many present and future stress factors. To address these questions, the originality of the present essay is to recognize the key role played by the concept of "stress", which is the reaction of a system to some factors that tends to perturb it from a reference state. The existence of stress leads to three possible types of characteristics for a system:

i fragility (system is prone to disability of its functions or even to destruction),

ii robustness or resilience (system is able to recover from not-too-large stresses), and

iii adaptiveness and transformation, leading to phase changes, regime shifts, modified behaviors and even to drastic structural reorganizations such as in biological mutations.

In this framework, we examine in detail the claim that stress can be beneficial and show that it is subdued within the earlier and more general concept of "adaptive systems" according to which systems evolve endogenously in symbiosis with the so-called stressors. The other essential role of stress in the evolution of systems is to promote rare intermittent rapid speciations, such as in punctuated biological evolution. We show that the concept of "antifragility" recently introduced by Taleb (2011) describes the quality of some systems that are designed to profit from particular stressors that stress other systems and to which they are not sensitive themselves. But, these so-called "antifragile" systems also exhibit vulnerability with respect to other stressors that lie outside their tailored design. Many presented antifragile systems are also much less productive that their fragile or resilient counterparts, showing the importance of recognizing the defined objectives. Hence, we conclude that antifragility does not exist per se and that the concept is misleading.

The present essay provides a rigorous definition of stress in corresponding systems. We describe how to measure stress, how to delineate the possible responses to stressors and we spell out propositions towards more resilience and sustainability. We emphasize the need for specifying the goals of a given system, from which the process of resilience build-in can be defined. We distinguish between exogenous versus endogenous sources of stress, and delineate the corresponding appropriate responses. We outline the main ingredients of resilience in terms of (1) the need for continuous multi-variable measurements and diagnosis of endogenous instabilities, (2) diversification and heterogeneity, (3) decoupling, (4) incentives and motivations, and (5) last but not least the (obvious) role of individual strengths. In this respect, propositions for individual training towards resilience are articulated. The concept of "crisis flight simulators" is introduced to address the intrinsic human cognitive biases underlying the logic of failures and the illusion of control, based on the premise that it is only by "living" through scenarios and experiencing them that decision makers make progress. We also introduce the "time@risk" framework, whose goal is to provide continuous predictive updates on possible scenarios and their probabilistic weights, so that a culture of preparedness and adaptation can be promoted. These concepts are presented towards building up personal resilience, resilient societies and resilient financial systems.

2. Definitions of stress

Defining stress is the first step towards a full understanding of risks, fragility, robustness, resilience and the development of efficient risk management. The word "stress" is part of the common vocabulary. However, in view of the widespread misunderstanding and confusion, rigorous and precise definitions are required. Before formulating a general definition of stress, it is useful to present illustrations through examples offered by different scientific fields.

In physics and more specifically, in continuum mechanics, stress is defined as a measure of the internal forces acting within a deformable body (Chen and Han, 2007). Quantitatively, we speak of a stress field defined as the ensemble of the stresses defined over all points within the body. Precisely, the stress at one point is a tensor that allows one to determine the force per unit surface that applies on any arbitrary fictitious plane specified by its orientation and going through that point. In a simple cylindrical geometry, an external force applied along the long axis translates within the body into a stress equal to the force divided by the area of the cylindrical cross-section. In equilibrium, the internal stresses sum up to balance exactly the external forces applied to the system. One can state the general result that the internal forces (and therefore stresses) are a reaction to external forces (stressors) applied on the body.

In biology, the endocrinologist pioneer, Hans Selye, introduced the concept of stress on the basis of his observations that many different types of substances and, more generally, perturbations applied to animals led to the same symptoms (Selye, 1973). The concept of stress in biology is thus based on the existence of non-specific responses of the body to the demands placed upon it. Transient perturbations, which do not exceed the natural regulatory capacity of the organism, lead to responses that ensure the resilience of homeostasis, the dynamical equilibrium characterizing living entities. In the presence of unrelieved stress, the body often transitions to pathological states associated with a change of homeostasis. This is analogous to the initial visco-elasto-plastic response of a mechanical system to an external stress, followed by creep that usually ends in the tertiary rupture regime (Nechad et al., 2005).

Common features can be observed in the interaction processes of different systems and their environments. Thus, the concept of stress was rediscovered, reused and often modified in various applied fields: organizational science (Cooper et al., 2001), seed science (Kranner et al., 2010), climate change and food security (Parry et al., 1999) and many other areas (Aldwin, 2007).

Several important characteristics of stress can be learnt from these examples:

1. stress is an internal response/reaction of a system to a perturbation called stressor (or stress-factor);
2. a stressor is a demand applied to the body that requires its reaction and adaptation;
3. a stressor elicits a non-specific response regardless of the nature of the stress, and even whether the stressor has a positive or negative consequences in the long term.

More generally, for biological as well as socio-economical systems, the non-specific response or "symptoms of stress" to a new demand involves increased:

i attention;
ii mobilization of resources;
iii concentration on key areas; and
iv recovery or exhaustion of the adaptive response and transition to pathological or crisis states.

In adaptive immune systems, (i) T and B lymphocytes first recognize the dangers, then (ii) mobilize the generating centers of antibodies that (iii) are finally directed towards and concentrated at the loci of insult. In social systems, the three first steps of the non-specific responses are typical of military-type intervention to cope with internal or external threats. In psychology, the first step (i) is associated with alarm, the second and third steps with resistance and the fourth step with exhaustion, as classified within the so-called general adaptation syndrome (Selye, 1973). More specifically, professionals facing acute situations, such as competitive pilots, athletes, surgeons and so on, go through the three first steps during their transient stressful activities. In economics, the response to economic difficulties is associated with (i) the characterization of the symptoms (solvency problems, budget deficit, increase of debt), (ii) the identification of reserves through expense cuts and reengineering of business and risk management processes and (iii) the reallocation of resources on key business lines or subsidizing. These measures may lead (iv) to a stabilization, or to a transition to a new favorable economic regime catalyzed by economic reforms and innovations or to bankruptcies in the context of firms, or to a disruptive transition to a new political order in the context of nations.

3. Measures of stress in social science

As a consequence of the complexity of social systems and the diversity of situations and applications, measuring stress in social sciences is a non-trivial issue. In contrast, in natural sciences, one often has the luxury of observing the stresses by their direct effects. In mechanics, direct measurements of stress within a system are often performed by observations of deformations of the body. In biology, the measurement of stress is obtained by observing the response of the biological processes to a stressor. However, in social sciences, the feedback loops as well as coupling mechanisms to exogenous factors are much less understood. As a result, the quantification of the stress level is performed indirectly via probabilistic approaches that introduce metrics of risks and/or resilience. These indirect ways of stress measurement in social sciences may be at the origin of the confusion in dealing with the concept of stress, incorrectly interpreted not as an internal response of the system to stressors but as the source of difficulties faced by the system.

3.1. Risk as measure of stress

Formally, risk is defined as the triplet of

1. a probability when available, or a level of uncertainty, or in the worst situation the formulation of the ambiguity corresponding to ask the question on the possibility for the occurrence of certain stressors;
2. a potential loss quantifying the possible impacts of the stressor;
3. a vulnerability and related counter-measures and mitigation techniques, that specify how disruptive the potential stressor to the system is
(Kaplan and Garrick, 1981).

The two first properties characterize the external forces or stressors that may influence the system. Together with the third property, which is specific to the system, they control the overall losses that the stressor can bring to the system. As a consequence, risk is understood as the combination of these three characteristics of the potential stressor. Thus, risk is equal or proportional to the possible internal response of the system, and therefore is a proxy for the stress developing within the system.

The simplest response of a system to a normal stress is non-specific and non-directional, which is comparable with the biological concept of kinesis. More resilient systems need to develop targeted reactions to stress, which is analogous to taxis in biology, defined as a directional response of a system to a stimulus or stimulus gradient intensity. In this sense, "stress taxis" can be defined as a response that, in the end, tends to unload stress off the system. For example, bacteria are wonderfully evolved organisms that demonstrate incredibly high resilience by using taxis and their corresponding simple behavioral rules.

3.2. Resilience as measures of stress

Resilience comes into several levels. The first two levels of resilience can be conveniently classified by using the theory of dynamical systems.

First level of resilience: Resilience is often defined as the speed of return to equilibrium (or more generally to the attractor characterizing the system) following a perturbation (Pimm and Lawton, 1977). Technically, the first level of resilience is referred to as "engineering resilience", which is a local concept. Engineering resilience is described by a local analysis, in terms of the stability of the linearized dynamics in the neighborhood of the equilibrium point. Indeed, resilience in this sense refers first to the stability of the equilibrium state, which occurs when all Lyapunov exponents are negative. Then, the speed of return to the equilibrium point is controlled by the largest (negative) Lyapunov exponent (i.e., the smallest one in absolute value).

Second level of resilience: In contrast, "ecological resilience" encompasses and generalizes engineering resilience by referring to the non-local dynamics occurring within the basin of attraction of the equilibrium state, defined as the set of initial conditions of the system that converge to that equilibrium state. While engineering resilience is a local concept quantifying the response of the system to small perturbations, ecological resilience describes the fact that a system state will return to its initial equilibrium as long as the perturbations remain within the basin of attraction of the equilibrium point, thus embodying non-local finite size perturbations that can be as large as the size of the basin of attraction itself, but not larger.

Walker et al. (2004) review four main components of ecological resilience of a system in its capacity to absorb disturbance and reorganize itself in order to retain essentially the same function. Using the dynamical system analogy with attractors and their basins of attraction, these four components are:

i latitude (controlled by the size of the basin of attraction),
ii resistance (controlled by the height of the barriers between attractors),
iii precariousness (controlled by the current position of the system within the basin of attraction),
iv iv. and panarchy (controlled by the way the attractor structure and its basin may change as a function of the scale of description through cross-scale interactions) (Gunderson and Holling, 2002).

Extending the so-called resilience triangle approach (private communication of Wolfgang Kroger, ETH Zurich, see e.g. Bruneau et al., 2003; Chang and Shinozuka, 2004; Pant and Barker, 2012), one can simplify the picture offered by ecological resilience by introducing four variables characterizing the response of a system to an external shock. Considering the variable W0 corresponding to a reference capacity, wealth or production level just before the shock, we define

i the maximum loss (1-λ)W_0,
ii a characteristic time τ_1 of reaction to reach the bottom level λ W_0,
iii the level Λ W_0 recovered,
iv after the characteristic recovery time τ_2.

In this simplified formulation, the resilience of the system is captured by the quadruplet of parameters (λ,τ_1, Λ, τ_2). Note that Λ could be larger than 1, corresponding to the situation where the shock has long-term beneficial effects by increasing the overall performance above the initial baseline W_0. Some systems may be characterize by Λ being smaller than λ, in which case, after a first loss of performance over a first reaction time τ_1, the system degrades further over a possibly different time scale τ_2 to an even worse situation. We should also stress that the quadruplet (λ, τ_1, Λ, τ_2) may not be unique but depend on the severity and duration (as well as possibly other characteristics) of the shock, so as to reflect the nature and amplitude of possible cascades occurring within the system.

Third level of resilience: The concept of viability (Aubin, 1991; Deffuant and Gilbert, 2011) extends further the idea by focusing on the conditions that the system must obey to remain "viable", for instance functional or alive. These constraints may not in general map precisely onto the set of attractors of the dynamics or may not even be attainable by the natural evolution of the dynamics and therefore may require continuous external management and control.

Fourth level of resilience: The dynamical system analogy has however its limit if taken too rigidly, because it fails to account for the fact that many biological, ecological and social systems may actually adapt, evolve and even transform fundamentally under the influence of stressors (Walker et al., 2004). This requires the consideration of other levels of resilience, which takes into account the possibility for the system to adapt its constituents so as to influence its resilience. This may correspond to a deformation of the basin of attraction, the fusion of initially distinct basins and other topological transformations. More generally, the dynamical system may incorporate stochastic components, such as deterministic, quasi-periodic or even random deformations of the attractors due to the modulation of some control parameters, as long as the conditions of viability are respected. Then, the system keeps its identity, but in a broader sense, even redefining itself while still keeping its ability to cope with the stressors. Pushed to the extreme, the system may even transform itself into a completely different structure via its capacity to evolve, as described by the theory of complex adaptive systems (Holland, 1975; Kauffman, 1993).

These considerations underline that the concept of resilience is dependent on the time scale over which the stressors act. For short-lived disturbances compared with the characteristic time scales of reactions of the system, engineering and ecological resilience are the relevant levels of description. At intermediate time scales, the issue of viability dominates, pushing for adaptation and redefinition of goals and processes. At the longest time scales, transformations may occur that are similar to natural selection and Darwinist evolution of species, seen as a transformation in response to changing geological and climatic conditions. In the context of man-made and social systems, Darwinist evolution is also relevant to understand the dynamics of human enterprises (Hannan and Freeman, 1977; Hite and Hesterly, 2001). Real life situations are likely to involve an interplay between a continuum of different time scales and thus between the different levels of resilience.

3.3. Links between risk and resilience as complementary measures of stress

To summarize, risk and resilience are two complementary revelations of stress. On the one hand, risk provides a measure of the nature and amplitude of stressors, present and future. As a consequence, from risk measurements, one can infer the possible level of stress that may develop within the system. On the other hand, resilience characterizes the internal stress response within the system, quantified by the capacity of a system to cope with stressors and remain essentially the same. In other words, resilience is the amount of stress that a system can bear without a considerable transformation.

Risk and resilience are inter-connected in another way through the concept of vulnerability (Birkmann, 2006; Cutter et al., 2008). On the one hand, vulnerability is part of risk, as a quantification of the potential amount of losses that are specific to a given system. But this vulnerability depends on the structural and adaptive properties of the system that make it either more prone to losses or less vulnerable via better mitigation techniques. In this sense, vulnerability constructs a bridge between risk and resilience. The processes favoring resilience will tend to decrease vulnerability and vice-versa.

The duality of stress expression in risk versus resilience is also apparent in the different possible responses of the system to stressors. These responses can be classified into three main classes: (i) fight, (ii) fly and (iii) transform.

i "Fight" is the typical response under relatively small risk and large resilience, which are associated with "normal" stress developing within the system. The "fight" response can be characterized by negative feedback loops tending to stabilize the system around its previous state, such as in the homeostasis state of living biological entities.

ii In contrast, the "fly" response corresponds to systems where risks and resilience are at comparable levels so that there is significant hazard for the system. By avoidance strategies, or some adaptation without major transformations and/or improvement of management, resilience can be improved so that the stressors can be addressed in order to ensure the preservation of the system identity.

iii Finally, when risk is large and resilience is insufficient, "extreme" stress develops within the system. Other than its demise, its survival requires considerable transformations of the system itself via the activation of positive feedbacks that drives it towards a new state.

The rational response to the presence of risks (the potential stressors and corresponding stress of the system) would seem logically to strive for always increasing resilience (the stress that the system can bear). However, there is always a cost-benefit balance between two extremes, the laissez-faire attitude of no investment in resilience as one extreme, and extreme risk aversion leading to attempts to over-control at the other. Building up resilience requires indeed to increase reserves, develop excess capacity, construct alternative supply chains, ensure redundancy, as well as investing in continuous education and training. But modern optimizing firms and societies work with the just-in-time philosophy and the constraint of ever lowering costs. This is often an impediment for building up resilience, as many examples show (Sheffi, 2005). It is a general observation that management in social systems strives to optimize this cost-benefit conflict, however, with often limited or even disappointing results to show for. In contrast, it is remarkable that natural systems often tend to evolve, converge and operate close to states that exhibit such a balance. These states are referred to in the modern literature as "self-organized critical" (Bak, 1996) or "at the edge of chaos" (Kauffman, 1993). This describes the tendency for coupled entities that interact over many repetitive actions to function close to a bifurcation point separating states that are too stable, from other states that are too unstable. A typical example is the human brain, for which there is a growing consensus that it operates close to or even functions at a critical point (Chialvo, 2006; Levina et al., 2007; Meisel et al., 2012; Plenz, 2012), separating a sub-critical state from a supercritical one. In the critical state, the brain exhibits the largest possible reactivity to novel external stimuli while, at the same time, showing stability of memory and other functional properties. If the brain was in the subcritical state, it would learn less efficiently by being not malleable enough and would be too slow to react in crucial situations. If the brain was in the supercritical state, its neural network would fire too much and too often, oscillating between extreme activity and exhaustion. Such a pathological state is actually found in epileptic patients (Osorio et al., 2010). In natural and biological systems, there are in general strong negative feedback mechanisms to stabilize the system and poise it at an optimal point between costly increase of resilience and costly neglect of the looming risks (Scheffer, 2009). The balance corresponds to a merging of the two responses - "fight" and "fly" - so that the system may combine both negative feedback reactions as well as adaptation to remain at the "edge of chaos".

In social systems, there is a lot of lip service paid to the goal for managers and policy makers to obtain this kind of optimal state. Actually, there is often an illusion of control (Langer, 1975; Satinover and Sornette, 2007; 2011) that it is possible to remove most of the risks and obtain an ideal state of resilience. One argument for the insufficient resilience of social systems (Diamond, 2004) is that, due to their complexity, they have not had the time to evolve (Walker et al., 2004) by the forces of "natural selection". This may be a part of the truth. However, we note that, for some social systems such as financial markets, there is ample evidence of an absence of convergence towards a stable dynamics, but rather the existence of persistent cycles of bubbles and bursts (Kindleberger, 2005; Sornette, 2003), notwithstanding experiencing many crises that, one would surmise, would have enabled investors to learn and avoid the next one (Reinhart and Rogoff, 2011). One possible explanation can be found in the incentives of investors to maximize their return on short time scales, leading to recurrent instabilities (Minsky, 2008). More generally, in many social systems, there is the ubiquitous problem that the short-term incentives are often not aligned with the long-term ones. This is associated with hyperbolic discounting (Laibson, 1998), which describes the general exaggerated preference for smaller immediate rather than larger delayed gratifications. Similarly, the incentives at the individual agent level are often incompatible with those at the society level, leading to social dilemmas (Kerr, 1983). It is also associated with the so-called public good problem and the problem of fostering social cooperation in particular in the context of socio-ecological systems (Ostrom, 1990). The rest of this essay aims at characterizing the conditions for breaking these kinds of stalemate.

4. Can stress be beneficial?

When thinking about stress, a first attitude is to find ways of reducing it or, when not possible, of developing passive and/or active defenses. But, there is a growing recognition that moderate levels of stress may be actually beneficial, both for health and for performance (Weiten and Lloyd, 2005; Hosenpud and Greenberg, 2006; Ritsner, 2010; Contrada and Baum, 2010). Is stress really beneficial per se?

4.1. System-stressor co-evolution under normal stress

For passive systems, stress is in general destructive, as in creep of materials where microscopic tiny damage events accumulate and lead to global rupture. In contrast, active systems can detect stress and use it as a guiding signal on the way towards better fitness to novel conditions. Thus, random or intended stressors are usable for the

i identification of the characteristics of stress by listening and analyzing reactions of the system to perturbations;
ii measurement of stress: (a) risks (observation of event probabilities, losses, vulnerability of the system) and (b) resilience ("exploration" of the stability landscape characterized by its latitude, resistance, precariousness and panarchy);
iii catalysis of learning, which promotes changes occurring through feedback mechanisms by adaptation towards better fitness under changing conditions, and of selection of specific features and implementation of contra-measures;
iv excitation of the system readiness, maintaining an engaged, interested and concerned state (in the spirit of the Soviet Union pioneer's motto "Always Ready!").

In section 2, we identified that symptoms of stress in a system include attention, mobilization of resources, concentrations on key areas, and so on. This may be viewed as positive consequences of stress for the function of the system. But, these changes are actually occurring at a cost, in particular that of a loss of resilience because the allocation of resources to cope with the stressor makes the system more vulnerable to other stressors. Thus, the optimization to cope with a first stressor should not be seen necessarily as a benefit of the stress. In general, optimization processes and coping with stress (or strengthening resilience) should be disconnected.

We also need to mention the cases in which some stress can be caused by a "positive" stress-factor (termed "Eustress" by Selye (1973)). For example, an eustress could be an economic reform that, after a period of adaptation, would lead to increased economic growth. Or, an extraordinary good news (learning about the return of a lost one or winning a huge lottery sum) may induce strong stress in the person. Again, it is not stress itself that is beneficial. Stress is a signal of a change of conditions and is a "guide" on the way towards adaptation or transformation to better fit to the new conditions, so that a system can survive and benefit from them.

Many situations where stress is argued to be beneficial, which we are going to cover at least partially in the following, follow the same archetype in which the system under consideration has co-evolved with the stress. In other words, the system is within an environment in which stress is unavoidable. Stress seems to be beneficial simply because the raison d'être of the system or of some of its key properties is precisely to cope and live with the ambient stress. Therefore, it is almost a tautology to find that the system needs stress or benefits from stress because it becomes dysfunctional if one of its main inputs, stress, is absent. We can therefore state that stress, at least up to a certain level smaller than the system resilience, is part of the normal system function and we refer to this situation as "normal stress".

4.2. Adapted systems co-evolved with their stressors

In this section, we provide several examples illustrating the concept that so-called beneficial stress occurs when the system under consideration has co-evolved with the stress.

4.2.1. Mammal immune systems, bones and muscles

Biology and medicine have probably been the first disciplines to recognize the co-evolved nature of stressors and of the stresses that develop within living systems. The immune system of mammals, in particular, provides arguably the best example illustrating what could be referred to with perhaps some exaggeration as a symbiosis between stressors (antigens) and system (antibodies). We underline that the example of the immune system provides a particularly important illustration, since its main role is indeed to defend the organism against disruptive intrusions by pathogens, in particular, which would like to exploit the organism for their own propagation. Consider first other types of homeostasis control processes in which the target variables are kept in a narrow optimal range with small fluctuations. This describes the "stable" homeostasis control for the regulation of the amounts of water and minerals by osmoregulation in the kidneys, the removal of metabolic waste by excretory organs such as the kidneys and lungs, the regulation of body temperature, the regulation of blood glucose level by the liver and the insulin secreted by the pancreas, and so on. In contrast, "The (immune) system never settles down to a steady-state, but rather, constantly changes with local flare ups and storms, and with periods of relative quiescence" as quoted in (Perelson, 2002), and see also (Perelson and Weisbuch, 1997; Nelson and Perelson, 2002). These flares can be understood as transient nonlinear reactions to fluctuating exogenous stressors as well as to expressions of the internal stress states. A growing body of literature indeed suggests that the incessant "attacks" by antigens of many different forms have forced the immune system to develop continuing fight and adaptation processes to ensure the integrity of the body (see Sornette et al., 2009b for a review and mathematical modeling). In this vein, the 'hygiene hypothesis' (Schaub et al., 2006) states that modern medicine and sanitation may give rise to an under-stimulated and subsequently overactive immune system that is responsible for high incidences of immune-related ailments such as allergy and autoimmune diseases. In this view, infections and unhygienic contact may confer protection against the development of allergic illnesses. For instance, Bollinger et al. (2007) suggested that the hygiene hypothesis may explain the increased rate of appendicitis (~6% incidence) in industrialized countries, in relation to the important immune-related function of the appendix. Sornette et al. (2009) concluded that, if the regulatory immune system was not continuously subjected to stressors, its adaptive component would decay in part and the defense would go down, thus letting the organism becoming vulnerable to future bursts of pathogen fluxes. They developed a mathematical model that demonstrates that the correct point of reference is not a microbe-free body (no stressors), but a highly dynamical homeostatic immune system within a homeostatic body under the impact of fluxes of pathogens and of other stressors (which include microorganisms such as bacteria, viruses, fungi, parasites, environmental load, over-work, overeating and other excesses, psychological and emotional factors such as anger, fear, sadness, and so on). The situation is analogous to the maintenance of healthy bones and muscles of a human being. For astronauts under zero-gravity (no weight stressor), loss of bone and muscle, cardiovascular deconditioning, loss of red blood cells and plasma, possible compromise of the immune system, and finally, an inappropriate interpretation of otolith system signals all occur, with no appropriate counter-measures yet known (Young, 1999). In other words, for bones and muscles, stress (in the real mechanical sense of the term!) is needed to avoid degenerescence and ensure appropriate strength in cases of need. In all these examples, stress is beneficial only because the systems are fundamentally defined in their aims and properties by their interactions with stressors. Biological evolution has weaved a complex network of interacting feedback loops that entangled fundamentally the systems with their stressors, making the later necessary for the normal function of the former.

4.2.2. *Human cooperation, competition and risk taking*

An enormous body of anthropological and ethnographic literature demonstrates that the level of cooperation between humans is exceptional both in quality and quantity (Henrich and Henrich, 2007), which explains the remarkable success of this single mammal species that nowadays controls a major part of the whole output of planet Earth (Steffen et al., 2004). However, the origin of this cooperation is still quoted as one of the 25 most compelling puzzles that science is facing today (Siegfried, 2005). Many mechanisms and contextual factors have been proposed to explain the remarkable level of pro-social behavior and cooperation between humans, such as kin selection, inclusive fitness, reciprocity, network reciprocity, group-level and multi-level selection, other-regarding preferences, relative income preferences, envy, inequality aversion and altruism (Axelrod and Hamilton, 1981). Two essential ingredients emerge: (i) the presence of differences in skills, contributions, rewards and retributions among group members and (ii) how perceptions and preferences drive human decisions and actions. In other words, not only exogenous stressors resulting from the environment such as predators but also within-group stressors have been found essential to promote cooperation. This has led to a significantly higher survival efficiency and larger fitness both for the group and for the individuals. Using agent-based models and analytical theory, Hetzer and Sornette (2011; 2012) in particular have shown that cooperation evolves at the level documented for humans only under two conditions: (i) agents exhibit disadvantageous inequity aversion, which is found to be evolutionary dominant and stable in a heterogeneous population of agents endowed initially only with purely self-regarding preferences; (ii) groups are "stressed" by random perturbations in the form of strangers migrating between co-evolving groups and who introduce different cooperation levels than those that would emerge from the group consensus in absence of the random perturbations. The underlying mechanism is related to the Parrondo effect describing situations where losing strategies or deleterious effects can combine to win (Harmer and Abbott, 2002; Abbott, 2002). Here, the random behavior is rooted in the exchange between groups and the asymmetry is inscribed in the punishment rule driven by disadvantageous inequity aversion. This constitutes a telling example illustrating that stressors have selected for enhanced cooperation via higher survival rates for groups and individuals. This became possible when cognitive abilities in our homo ancestors increased sufficiently to allow the exploitation of this new "resource" of enhanced cooperation beyond that observed for our primate cousins, again illustrating the co-evolution between stressors and system's abilities.

Another important characteristic of humans is that high male-male competition for reproductive success has been permeating the history of modern humans (200'000 years ago to recent times) and has contributed through gene-culture coevolution to create gender competitiveness-related differences. Favre and Sornette (2012) have recently introduced a simple agent-based model that explains the high level of male-male competition and risk taking as rooted in the unequal biological costs of reproduction between males and females. This cost asymmetry has promoted females' choosy selection of alpha-males who have better chance to propagate genes via the natural selection of the fittest (Baumeister, 2010; Ogas and Gaddami). This causes male-male competition and male's arm race for signaling their qualities, which takes the form of stronger risk-taking behavior (Diamond, 2002). This further cascades into higher male than female death rates through risky signaling and results in a smaller male than female effective breeding population, both because females select a subset of males for reproduction and because of male's higher death rate. Remarkably, this mechanism can be checked quantitatively through its prediction for the ratio of the Time To the Most Recent Common Ancestor (TMRCA) based on human mitochondrial DNA (mtDNA), i.e. female-to-female transmitted, which is estimated to be twice that based on the non-recombining part of the Y chromosome (NRY), i.e. male-to-male transmitted. It appears that we are all descended from males who were successful in a highly competitive context, while females were facing a much weaker female-female competition. Stresses have appeared endogenously in the human population as a response to the unequal biological costs of reproduction (itself a stressor), leading to males' arm race in risk taking (another set of stressors) and cascaded into extraordinary implications for the development of the human species and its conquer of the world (Baumeister, 2010). One can argue that the high level of risk taking of human males have been beneficial for mankind, through the exploration of unknown territories and the development of inventions, in the end making stressors, via enhanced risk-taking by males, the engine of progress. The causal flow "reproduction inequality ⇒ female strategy ⇒ male risk taking" of stressors can thus be seen as an intrinsic part of the making of mankind, providing another example of the entangled nature of the human system and its stressors, the latter being beneficial on the long term as a result of their coexistence and co-evolution. Pushing this reasoning, one can thus conclude that being human is to use one's superior cognitive abilities to take risks beyond the biological laws that enslave other animals.

4.3. Change of regimes under extreme stress

Nature and human societies exhibit many cases in history and in recent times when stress surpasses the resilience level of the system. We refer to such response of the system as "extreme stress" because of dramatic consequences it may lead to. Sources of extreme stress can be tracked using the measures of stress that were described above - risk and resilience - and include:

i extreme possible stressors that are characterized by low probability and/or huge losses, for example, very rare events of enormous impact or previously unknown events (black swans (Taleb, 2007));

ii unbearable stress that the system is not capable of coping with, showing extreme vulnerabilities (for example, disfunction of critical systems) and/or zero resilience, when even a tiny perturbation can lead to a change of regime. Examples of such systems include those (1) optimized to the edge of maximum efficiency, such as the just-in-time Toyota supply chain and inventory management system and (2) close to a tipping point due to developed endogenous instabilities, leading to dragon-kings (Sornette and Ouillon, 2012).

In the worst cases, this leads to the death or demise of the corresponding organism or system, as for instance documented by J. Diamond (2004) for human societies. In other situations, the system evolves to another regime, in which different properties that were dormant come into play or novel ones are forced to evolve for the survival and success of the system. The following two subsections examine a number of real life examples illustrating the occurrence of regime shifts and evolution under extreme stress

4.3.1. Biological and other transitions

The existence of changes of states promoted by extreme conditions is perhaps best incarnated by biological evolution. Contrarily to the initial view held by Darwin that evolution is generally smooth and continuous, occurring by the cumulative effect of gradual transformations, the theory of punctuated equilibrium in evolutionary biology describes the evolution of species as a sequence of stable states punctuated by rare and rapid events of branching speciations occurring under the stresses resulting from climatic, geographic and other possible evolutionary stressors (Gould and Eldredge, 1993). Since its introduction (Eldredge and Gould, 1972), this theory has received strong empirical support (Gould, 2002; Lyne and Howe, 2007). It holds that most species exhibit little evolutionary change for most of their geological history, being adapted to their niches. But, something happens, such as an extreme disturbance, that pushes the species to branch into novel species, often with the demise or altogether change of the original species.

Many scientists view the abrupt changes occurring in the sequence of punctuated equilibria as due to catastrophic causes, such as the famous Chicxulub asteroid (Schulte et al., 2010) or enormous volcanic eruptions in the so-called Deccan trap epoch (Courtillot and McClinton, 2002), or both (Archibald et al., 2010) ending the reign of the mighty dinosaurs about 65 million years ago. Starting with Bak and Sneppen (1993), others have argued for an endogenous origin, using the analogy with the concept of self-organized criticality (Bak and Paczuski, 1995; Bak, 1996; Jensen, 1998; Sornette, 2004, chapter 15). According to complex system theory, out-of-equilibrium slowly driven systems with threshold dynamics relax through a hierarchy of avalanches of all sizes. Accordingly, extreme events can also be endogenous.

The exogenous versus endogenous explanations may actually represent two complementary view points since, in reality, they are often entangled. Indeed, how can one assert with 100% confidence that a given extreme event is really due to an endogenous self-organization of the system, rather than to the response to an external shock? Most natural and social systems are indeed continuously subjected to external stimulations, noises, shocks, stress, forces and so on, which can widely vary in amplitude. It is thus not clear a priori if a given large event is due to a strong exogenous shock, to the internal dynamics of the system, or maybe to a combination of both. Sornette et al. have advanced the hypothesis that specific dynamical signatures of precursors occurring before and relaxations following extreme events lead to a classification of possible regimes and the possibility to resolve the endo-exo conundrum. This applies broadly to many complex systems (Sornette and Helmstetter, 2003; Sornette, 2005), for which it is fundamental to understand the relative importance of self-organization versus external forcing, as documented for financial shocks (Sornette et al., 2003), commercial sales (Sornette et al., 2004), and for the dynamics of fame of YouTube videos (Crane and Sornette, 2008). More generally, in addition to biological extinctions such as the Cretaceous/Tertiary KT boundary (meteorite versus extreme volcanic activity versus self-organized critical extinction cascades), this question applies to commercial successes (progressive reputation cascade versus the result of a well-orchestrated advertisement), immune system deficiencies (external viral/bacterial infections versus internal cascades of regulatory breakdowns), the aviation industry recession (9/11 versus structural endogenous problems), discoveries (serendipity versus the outcome of slow endogenous maturation processes), cognition and brain learning processes (role of external inputs versus internal self-organization and reinforcements) and recovery after wars (internally generated (civil wars) versus imported from the outside) and so on. In economics, endogeneity versus exogeneity has been hotly debated for decades. A prominent example is the theory of Schumpeter on the importance of technological discontinuities in economic history. Schumpeter (1942) argued that "evolution is lopsided, discontinuous, disharmonious by nature... studded with violent outbursts and catastrophes... more like a series of explosions than a gentle, though incessant, transformation."

4.3.2. Political and economic transitions

Consider the fall of the Berlin wall in October 1990 associated with a series of radical political changes in the Eastern Bloc. Over the period from 1989 to 1992, many east European countries engaged in a transition from a centrally planned economy to a democratic and market economy. Using agent-based model simulations and economic data, Yaari et al. (2008) discovered that all countries' GDP (gross domestic product) as well as other indicators of economic development (such as the number of privately owned enterprises) evolved through a generic J-curve, corresponding to a first phase of strong decay followed by a recovery and, for some countries, a transition to a growth rate surpassing significantly the levels under socialism before 1990. The first decay arch of the J-curve corresponds to the progressive demise of the "old centrally planned economy", whose shrinkage dominates the rise of the "new" free market economy (Novak et al., 2000). The second rising arch of the J-curve embodies the progressive transition to the "new economy" that burgeons as a response to novel conditions (Challet et al., 2009). In the case of Poland, Yaari et al. (2008) found that the new economy principally developed around a few singular "growth centers" associated with pre-existing higher education poles, which was followed by a diffusion process to the rest of the country. The centers of education were thus the main engines of the resilience and adaptability of the Polish nation to the new conditions. In contrast, other Eastern European nations, such as Ukraine or even Russia, have fared much less well (Guriev and Zhuravskaya, 2009): for them, the transition resulted in a long lasting economic crisis that only recently has started to show observable improvement.

Let us scrutinize the economic transition in Russia. For a decade since the Berlin wall event, Russian GDP has been declining, with continuing huge drops in output and high levels of inflation. Russia went through a Great Depression more severe than that in the U.S. in the 1930s, with a decline in industrial production of over 60% from 1992 to 1998 (vs. some 35% decline in the U.S. Great Depression from 1929 to 1933), leading among many woes to the destruction of agriculture, deteriorating social conditions, health, education, environment, law, science and technology, high inflation and the destruction of the middle-class which is often the guardian of, as well as condition for, a functioning democracy. The Russian economy has been characterized over this time period as being riddled with crime and corruption. The transition was not to a market economy but rather to a criminalized economy, where the criminals established their own institutions in a process of self-organization (Intriligator, 1998). The reasons for these problems have been identified (Intriligator, 1997; 1998): by endorsing a stabilization program of the Russian economy based on liberalization of prices and the privatization of enterprises, the Yeltsin administration neglected the well-known but often forgotten fact that free markets require strong institutions, and in particular a legal system, courts, lawyers, law enforcement; property rights, and so on, so that business contracts are enforced rather than subjected to the whim of the strongest. Moreover, a strong government is at the core of market economies, as shown by numerous anthropological and historical studies documented for instance in Graeber (2011). Russia's transition illustrates that externally imposed conditions, fundamental internal situations as well as a badly chosen design of governance (without institutions and working legal system) led to a new regime that has struggled for a very long time to recover and establish a functional state for the well-being of the people (Guriev and Zhuravskaya, 2009).

The so-called Arab spring that began in Dec. 2010 constitutes another telling illustration of our thesis. This revolutionary wave of demonstrations and protests occurred in the Arab world, leading to the ousting of the leaders of Tunisia, Egypt, Libya and Yemen and civil uprising in other neighboring countries. While media reports and scholars have often viewed the Arab Spring movements as positive steps towards more democratic governance, some skepticism is in order when examining the post-Gaddafi outcome in Libya for instance. Research at the NECSI suggests persuasively that the triggering factor for many if not most of the upheaval movements observed in arab as well as other poor countries around the world coincide with rapid and large rises of food prices (Lagi et al., 2011; Bertrand et al, 2012). Indeed, commodity prices more than doubled in 2008 due to a combination of environmental factors, the accelerating needs of booming countries such as China as well as speculation (Sornette et al., 2009a). As a consequence, world food prices skyrocketed, making many households' subsistence reach a crisis level. The inability of the governments of the concerned countries to cope with these stressors led to the transitions (or in many other cases to the search for the resolution of quite unstable states) to what can still be seen as evolving situations in search of an equilibrium. Whether the outcomes in Libya or Egypt are positive remains to be determined as the region has become very unstable and the future remains highly stressful and uncertain for most of the population.

This is reminiscent of the French revolution of 1789: more than enlightenment ideals, economic factors arguably played indeed a crucial role. As a result of bad harvest over most of the decade preceding 1789, a large part of the French population was exposed to strongly rising bread prices (the main food), leading to hunger and malnutrition. In the absence of adequate reactions by the government to the climate stresses that were adding to a very large national debt and an antiquated tax system weighting unfairly on the working class, the resulting discontent population became prone to push for major changes that culminated with the storming of the Bastille. Similarly to the situations resulting from the Arab spring, one should be cautious to claim that the extraordinary changes resulting from the food price stressors (among others) have always and systematically been for the better in all dimensions. The situation is perhaps best captured by the apocryphal statement of Chinese premier Zhou Enlai during President Richard Nixon's visit to China in February 1972: "too early to say" when referring to the assessment of the implications of the French revolution (he was in fact probably referring to the turmoil in France in May 1968 (Campbell, 2011)). Notice also that there is clear evidence that the French revolution has led to much bloodier wars in which whole nations have become involved in large scale conflicts involving many casualties (Cederman et al., 2011), showing again the relativity of the values of the regime shifts and their often unintended consequences.

These examples have illustrated two main points:

i the ubiquity of (rare) regime shifts due to the combination of abnormally large external circumstances (that are bound to occur in any nonlinear system if one waits long enough) and internal facilitating processes limiting the build-up of adequate resilience;

ii the value (in terms of economic consequences, change of well-being, moral level, culture) of regime shifts is open to debate, depends on the time horizon (beneficial short-term but detrimental long-term, or vice-versa) and is arguably relative.

All the examples treated in this subsection refer to situations in which scholars and observers would rate the pre-existing regimes as (to various degrees) undemocratic, oppressive and in opposition with the enlightenment ideals. As we shall elaborate in section 5 on recipes for resilience, much of the strength of a nation rests on the cohesion between its citizens that is called upon at times of stresses. In this respect, Arab countries, the countries of the Soviet Bloc, and France under the Bourbon dynasty developed modes of governance that embodied the roots of their demise, such as increased inequity and rigidities. One should not develop however the impression that this situation is a unique attribute of countries that do not embrace the modern western version of market economy and of democracy (which, by the way, is not a unique governance process of course but comes in many kinds and degrees).

Consider the situation of the largest western economies, including the United States of America, Japan and Western Europe, whose indebtedness have reached, according to many analysts and pundits, unsustainable levels (Reinhardt and Rogoff, 2011). Scenarios for the next decades encompass the possibility for global critical transitions at worst or, at least, the need for massive readjustment of expectations (which is a polite way to say that retirees will get much less and after working significantly longer, average social coverage will shrink much further, standard of livings will at best plateau with many signs of deterioration for the median household). Here again, one can argue that the western economic systems have been built on a model of run-away indebtedness that, on the "short term" of the past several decades, brought extraordinary gains, at the cost of increasing systemic and global risks (Sornette and Woodard, 2010). The on-going crisis of debt-strangled European nations is far from finished, as nothing has been done in depth to address the problems of insufficient growth of productivity and innovations (Sornette, 2010), of the demographic bottleneck, and of reigning on wasteful over-spending beyond one's means by addicted consumers as well as nations spoiled by the failure of democracy replaced by demagogic politics (Gore, 2007). The US should not be forgotten either, if only because its financial system is effectively bankrupt, but held artificially alive by rounds of buying toxic assets by the Federal Reserve and the successive spells of so-called quantitative easing. An even greater crisis if possible is probably awaiting Japan, which relies on the policy of essentially zero-interest rate in order to cope with a total debt that dwarfs that of all other nations. The policy of ultra-low interest rate seems to become the new reference point of debt-strangled nations in order to be able to honor their interest payments, which yet not fully appreciated consequences concerning the transfer of wealth between generations and the possibility to face the huge retirement liabilities. Globally, the diagnosis is clear: these systems have built economic organizations that contain in themselves the seeds for monstrous systemic instabilities towards major re-organizations. The 2008 US crisis and the 2010-2012 sovereign European debt crisis are probably nothing but the premises of much more significant crises at the global scale. Such a prediction is warranted on the observation that none of the real causes of the crises have been addressed and only superficial short-term remedies have been offered until now (Mauldin and Tepper, 2011; see also chapter 10 of Sornette (2003) which is based on Johansen and Sornette (2001) and, more recently Akaev et al. (2012)).

Thus, we can add to the two points (i) and (ii) above a third one:

iii social and political systems seem to be intrinsically unstable on the long term, building up internally the mechanisms of increasing vulnerabilities via the very processes that seem initially the most favorable.

Resilience is therefore a fundamental question that needs to take into account both the conflicts between time scales (generations) and the unintended consequences of short-term innovations and improvements (Ferguson, 2011).

4.4. Debunking "anti-fragility"

It is appropriate to end the present section, discussing whether stress can be beneficial, by the extreme view proposed by Taleb (2012) summarized under the vocable "antifragility". According to this concept, "antifragile" systems may not only resist and recover efficiently from stressful events but may actually benefit from them in very direct ways and on the short term. Taleb lists a number of examples illustrating this view: muscles and bones, owning insurance or financial derivatives, decentralized organization and so on. If correct, the antifragile concept would contradict our whole construction presented above. To understand the source of disagreement, we now dissect Taleb's proposition. In a nutshell, antifragility describes the quality of some systems that are designed to profit from particular stressors that produce stress in other systems and to which they are not sensitive themselves. But, as we are going to show, these so-called "antifragile" systems have also their own vulnerability to other stressors that lie outside their tailored design.

4.4.1. The put option paradigm

The example that captures the essence of the whole "antifragility" argument is that of financial derivatives. Consider specifically a put (also called "sell") option written on some underlying financial asset. The later has values that fluctuate more or less randomly, with sometimes large excursions in the positive (gains) as well as negative (losses) ranges. An investor owning this asset will be exposed to possible rare large losses, the so-called tail events. The investor's investment is thus a priori vulnerable to the occurrence of financial shocks that may hit his asset and make it fall abruptly. Fragility is particularly acute if, as such time, the investor needs to cash out for some consumption needs (unforeseen medical expenses or student university tuition for his children) at the much lower asset value following the crash. Another investor, who has bought a put option of that same asset, has a diametrically opposed perception of the situation: when the asset plunges, the value of his put option sky-rockets upwards. In the terminology of antifragility, the put option investment of the second investor is antifragile, since it profits from large negative price movements that hurt most other investors. The put option paradigm is actually underpinning the whole antifragility concept when applied to general situations, as developed in Taleb and Douady (2012). To summarize, Taleb advocates strategies and policies that construct effectively put options everywhere!

Let us clarify how a put option works. First, it needs a risky asset or a basket of risky assets that are subjected to the influence of many natural and social factors so that its value fluctuates with sometimes large amplitudes. Second, it needs a counter party, say a bank, which accepts to create the put option and sell it to the second investor. In the case when the put option is exercised, the counter party has to pay for the gain of the option owner. The put option strategy is thus conditional on others taking the other side of the risks.

It is important to realize that the put option strategy is built on the premise that it can only work when endorsed by a minority of investors, at the expanses of the others. Take the example of the so-called "portfolio insurance" strategy developed in the 1980s by Leland and Rubinstein. Large institutional investors wanted to insure their large portfolios against possible drops of the stock market. For this, the simplest and most efficient strategy consists in buying put options on the assets held in the portfolios. However, the sheer volume of put options needed was beyond what banks and other option writers would be able or willing to offer. Or, if offered, the requested prices would have been prohibitive. Leland and Rubinstein then used the replicating construction of the Black and Scholes option pricing formula to devise a simple and effective way of constructing synthetic put options just based on the underlying assets and on bonds. The synthetic put options thus created led to a flourishing business where, at the time just before the crash of October 1987, more than one third of all US institutional investors had implemented the Leland-Rubinstein so-called insurance portfolio strategy (MacKenzie, 2008). The weakness of this whole construction however was revealed as markets started to stumble the week before "Black Monday" 19 October 1987. Because the synthetic put options operate by selling the underlying stocks when the later decreases in value, as the stock values start to go down, the synthetic put option strategy led to sells, pushing prices further down, these losses aggravating the negative sentiments of the markets, leading to an avalanche of sells reinforced by the technical implementation of the synthetic put options leading to a vicious positive feedback to the bottom. After the crash of October 1987, many pundits and scholars have concluded that, with a large probability, synthetic put option strategies were responsible for aggravating strongly the severity of the crash (Barro et al., 1989). What was supposed to be a bullet-proof strategy turned out as a catastrophe due to its hidden vulnerability with respect to synchronization. In other words, buying put options works when you are in the minority and no collective herding behavior occurs. More generally, the whole business of insurance is based on diversification of exposures. This message was vividly brought home to major insurance and re-insurance companies in the aftermath of the 9/11, when the capital stored in stock markets needed to be sold to compensate clients for their losses plummeted at the same time. This illustrated another mechanism of fragility of the supposed antifragile insurance strategy.

The 15 September 2008 Lehman Brothers bankruptcy and 16 September 2008 AIG official bail out demonstrate another fundamental fragility of the antifragile put option strategy. In short, major investment banks around the world had invested in CDO (collateralized debt obligations), which are securitizations of mortgages offered to millions of American households. Many of these investment institutions search for ways to insure their exposition to possible losses on the CDOs by buying massive amounts of CDS (credit default swaps) from counterparts, the most famous and by far largest being AIG, the then largest insurance company in the World. Different from what their name suggests, CDS work essentially as put options paying large amounts when the underlying CDO losses value and/or when some trenches of the CDOs start to default. Buying CDS was a perfect antifragile strategy to profit from the rather visible problems looming as a result of the enormous real-estate bubble that has developed in the US from the early 2000s to 2007. Except for one thing: the credit risk of AIG was not considered. Default of AIG was inconceivable. The problem is that the collective use of the antifragile CDS strategy led to such an enormous exposition of AIG to a downturn of the US real estate market that its total capital base became insufficient, finally leading to its quasi-bankruptcy and its final salvation by a massive injection of capital from the US treasury and a consortium of investment banks. The so-called antifragile CDS strategy backfired to systemic proportions, whose real consequences are still to be solved at the time of writing. Moreover, for an inner circle of investment banks, the CDS strategy turned out to be really profitable, though not from the intrinsic structure of the strategy but from playing the fear to the public of a global financial and economic meltdown as well as from using high-level political connections. The bail-out packages, which were put in place in September 2008 and following months, ensured the payments of most of the liabilities at 100% face value (which AIG could not longer support) to the major investment banks. The weight of these payments was in the end supported by the taxpayers.

In sum, these dramatic examples illustrate that antifragility does not exist. In general, for systems subjected to variability, noise, shocks and other random perturbations, it is possible to develop strategies that, on average, benefit from variability, but not any variability. Such strategies are designed to profit from the variability of particular stressors. Simultaneously, they are vulnerable to other stressors. The refusal to accept this fundamental characteristic (or intrinsic weakness) shared by any strategy or system is very dangerous, as it may lead to unexpected shocks or intended manipulations by insiders. For instance, in the financial sphere, antifragility is a name for the exploitation of a situation that turns losses for most into gains for some by special design, which is, however, vulnerable to non-anticipated occurrences. Moreover, the so-called antifragile strategy can contain the germs for large externalities, leading to systemic crises for which neither the strategy itself nor the system are prepared for.

4.4.2. Can antifragility be beneficial itself?

Taleb (2012) has provided many tentative examples of supposedly antifragile systems, putting them in contrast with fragile and robust systems. For each instance (i-vii) below, the antifragile system (according to Taleb) is indicated in boldface and contrasted with its opposite fragile version:

i civilization (*nomadic and hunter-gatherer tribes* versus post-agriculture modern urbanization);
ii production (*artisans* versus industry),
iii science/technology research (*stochastic tinkering* versus directed research);
iv nature of the political systems (*decentralized political systems* versus centralized nation-states);
v decision making (*convex heuristics* versus model-based probabilistic approach);
vi literature (*oral tradition* versus books and e-readers);
vii reputation (*artists or writers* versus academics, executives and politicians) and so on.

In all these examples, one notices that the antifragile system is much less productive than its fragile counterpart. In example (i), the capacity to support larger and growing populations has received an enormous boost with the introduction of agriculture while hunter-gatherer tribes had zero or very small growth. A typical North American family now commands a quantity of artifacts equivalent to or larger than that of a pharaoh at the peak of the classical pharaonic civilization. This illustrates that, in example (ii), the elaborate supply chains of modern industry based on the collaboration between millions of workers delivers enormously more than the whole summed contribution of individualistic generalists. In example (iii), the classical Greek tradition let place after many centuries of “stochastic tinkering” to an organized scientific production in the last few decades that dwarfs absolutely the knowledge accumulated earlier. In example (iv), nation-states have been able to mobilize resources unheard of decentralized political systems. Clausewitz (1984) [1832] in his classic book “On war” observed that the French revolution introduced the nation state, which led to global wars with enormously more resources, an hypothesis recently supported quantitatively using statistical comparative history (Cederman et al., 2011). In example (v), heuristics may often work for simple everyday problems and when immediate quick-and-dirty solutions are required, but would be unreasonable for decision making and management in sophistical modern systems dealt with by surgeons, airline pilots or technicians of nuclear plants. In the case of literature (vi), it is clear that oral tradition would not fail if electricity is no more available but, on the other hand, it is a very inefficient and low-density information medium, quite unsuitable to share and store the explosive amount of modern knowledge. Lastly (vii), academics, executives or politicians have developed extraordinary specialized skills that are (in principle) translated into positive reputation. A positive reputation serves the goal of producing more or delivering higher quality services and/or of being trusted. In contrast, some artists and writers just need any type of reputation as long as people and media speak about them, because their business is in a sense to bank on their fame. Pushing Taleb's reasoning to the extreme, one could conclude that being a beggar is one of the most desirable antifragile state to be in, since the person has nothing to lose and can only benefit (if he survives) from any change of his position. The condition "if he survives" actually demonstrates the essential hidden assumption underlying antifragile examples. Otherwise, as soon as there is something to loose, to disproof or the possibility of a disfunction, as when owning assets, possessing a reputation, using a decision model, or production scheme, there are many additional stressors that could cripple the system. Being rich, young, healthy, beautiful and loved is the ultimate fragile state, but who would exchange it for its absolute antifragile poor, aging, ill, ugly and lonely alter ego.

4.5. Can stress be beneficial? Our answer

To summarize, we have shown that stress is unavoidable and that systems co-evolve with their stressors. The survival of a system depends on its ability to cope with and adapt to numerous stressors. In this sense, the life-span of the adapting system is relatively longer than those of many of its stressors. These stressors, coming one after another, are progressively shaping the system, demonstrating sometimes a true symbiosis and an astonishing emergence of new features that can be beneficial for the system itself. In evolutionary biology, non-visible or "neutral" mutations occurring in the presence of internal stresses as well as small external stochastic perturbations, and which leave fitness unchanged, are considered beneficial because they improve the system's robustness (Kimura, 1983; Ciliberti et al., 2007). They provide a diversification by enlarging the toolbox of defense without disruption and prepare for major jumps when necessary or when ready (Wagner, 2005; Ciliberti et al., 2007). This concept seems to have broader applications, as recently proposed to quantify software robustness (Schulte et al., 2012). Finally, extreme stressors are relatively rare events, but they play an exceptional role in creating the global landscape and activating the mechanism of natural selection. Their magnificent power gave rise to legendary names - "dragon-kings" (Sornette, 2009; Sornette and Ouillon, 2012), for the extreme stressors of endogenous nature, and black swans (Taleb, 2007) that are characterized by exogenous sources.

The response of a system to stressors depends on the level of stress within it. To make the system more efficient and flexible, it is important to learn how to use normal stress as a signal of on-going changes and as a guide for needed adaptation to better fit to the evolving conditions, so that a system can survive and benefit from them. In the presence of extreme stress, resilience, that is, conservation of the status quo, may not be anymore an option and the resources should be directed towards an unavoidable transition to a new regime that can bear or even profit from the stress: in the words of Giuseppe Tomasi di Lamedusa, in 'The Leopard': "If we want everything to stay as it is, everything will have to change."

In Section 5, we propose strategic principles for system resilience and describe some of them in details. However, the adoption of strategic principles in most cases would require global systemic changes and would face numerous difficulties, partially described at the end of section 3.3. Therefore, in Section 6, we discuss some of these limitations and propose original operational solutions.

5. Recipes for resilience

5.1. Generic recipes for resilience

The systems that were previously mentioned are very different, and so are the conditions of their functioning and the stressors they face. Nevertheless, from the fact that stress is a non-specific response of a system that depend weakly on the type of stressor, it derives that the development of generic recipes to cope with stressors is both possible and crucial for strengthening its resilience.

We propose the following brief synthesis of strategic principles for the sustainable development of any system, which borrows from a variety of risk management thinkers, from Sun Tzu's "The art of war" (circa 500 BCE), Clausewitz' "On war" (1984) [1832], John Boyd's "certain to win" strategy and his OODA (observe-orient-decide-act) loop (Boyd, 1986; Richards, 2004) and Sheffi (2005). While rooted in ancient wisdom, their modern framing and phrasing do not diminish their reach and eternal relevance.

1. Develop strategic vision; orientation and focus on the present and future, and not on the past; establish *clear goals* (subsection 5.2),
2. build up, through investment and/or education, *fundamental values, right incentives and fair remuneration* (subsection 5.3),
3. diversify and promote *heterogeneity*, as well as *decoupling* of key components for sufficient redundancy,
4. develop operational mechanisms to *enforce contracts*,
5. promote *transparency, communication and ethics.*

At the operational level, tools for quantification of stress signals and learning from them should be put in practice in order to cope with stress effectively, i.e. to improve (i) the quality of decisions in the presence of risks and (ii) the management of resilience. These tools are to serve the following goals:

a development of individual strengths together with awareness of one's limits,
b promotion of collective action and collaborations,
c analysis and classification of stressors,
d risk identification and tracking,
e continuous measurements and diagnosis of endogenous instabilities,
f never ending verification and validation,
g always keeping on edge by questioning assumptions and existing processes.

This last point is easy to formulate on paper but much harder to implement in practice, if only because of the common adage that "No one sees any pressing need to ask hard questions about the source of profits, of success, or stability, when things are doing well." Building resilience requires indeed a kind of paranoic obsession that things could go wrong, when everything appears to be fine. Sections 5.3.3 and 5.4 provide concrete examples of such operational tools.

5.2. Formulation of goals and objectives

The first step on the way towards implementing the strategic principles for the sustainable development of a system is to identify and spell out the goals and objectives, which can also be called utility functions of the system. In this subsection, the strategies and methods of resilience growth are outlined into accordance with different types of goals.

1. At the most basic level, a first goal is to ensure survival, which calls for the measures promoting viability that are described in section 3.2, in particular using stress as information and being always ready for managerial actions to ensure that the system remains in its basin of attraction.
2. A second type of goals is often the conservation of the status quo, of existing wealth, of present standard of living. This triggers what we referred to as the "fight" response, which applies when the stress is significantly smaller than the existing resilience of the system. However, many systems, human societies and organizations in particular, reach high levels of wealth, which were obtained at the cost of strong optimization, decrease of reserves, indebtedness, increase of inter-dependencies (Diamond, 2004), which result in loss of resilience. In these situations, the fight response to maintain homeostasis at such high development levels is simply not possible in the middle and long term, because even small stressors will in the end be enough to trigger a change of regime due to the endogenous build-up of a critical fragility. As a vivid and painful example, one can argue that the present ongoing sovereign debt European crisis belongs to this class. Only with a profound reassessment of goals taking into account the realities of the globalized economy and the structural unbalances underlying the artificial construction of the euro dream, can one hope to address the systemic nature of the European conundrum.
3. A third type of goals, often observed in high-tech industries for instance, is for an entity to become and stay the leader among its pairs, hence developing highly competitive attitudes and strategies. IBM, Toyota and Apple are different examples of firms that were able to get to the top and remain there for longer than thought initially possible. For IBM, this was through its evolution from a mainframe computer hardware company to a service provider offering all possible integrated solutions to a large range of customers, thus redefining continuously what is the essence of being IBM. For Toyota, the empowerment of the factory workers, instructed to focus on the delivery of just-in-time products, led to a remarkably motivated and productive workforce delivering high quality products for more than 50 years. But the 2010 car recalls due to the sticking accelerator pedals and failing electronic throttle controls demonstrated that bureaucracy, overconfidence and weak management have lately underpinned Toyota's fall from grace. Apple's remarkable success can be attributed to its focus on innovation aimed at surprising and enthusing customers, by functioning as a secret organization with a self-perpetuating start-up culture. For these companies, resilience at the top requires internal engineering of their ever on-going mutation, aiming at shaping the future rather than reacting to it, in the spirit of "You don't wait for the future. You create it." (Hwang Chang Gyu, 2004).
4. In the modern world, the economic language and agenda dominates, with such concepts as utility function (assumed to capture people's goals) and growth of GDP (gross domestic product) taken as the universal measure of improvement and success. But, too little attention is given on what the US founders enshrined in the US constitution as one of the three main goals of well-functioning societies, namely the pursuit of happiness. In the United States and in many other industrialized countries, happiness is often equated with money. This simplifying assumption provides a convenient way of quantifying and comparing heterogeneous preferences of different agents within a unifying framework. This money (or economic utility function) approach has shaped our culture. Only the small Himalayan kingdom of Bhutan has made its priority to grow, not its GDP, but its GNH (gross national happiness). According to King Jigme Singye Wangchuck, Bhutan's goals are to ensure that prosperity is shared across society and that it is balanced against preserving cultural traditions, protecting the environment and maintaining a responsive government. In our context, this can also be interpreted as promoting a resilient society, based on (i) robustness anchored at the individual level (a happy and balanced person is arguably more robust in her behavioral response to stressors) and (ii) through cohesion within the society build on a common understanding that ethical behavior is fairly rewarded and equity (and not "equality" as in communism) is the standard reference.

The development of a strategy requires an out-of-the-box thinking and the consideration of multi-dimensional objectives. Setting up goals often crucially depend on the time scales of interest as well as on the size scales (individual versus group versus society). There are well-known differences in goals and welfare attained at the individual versus collective levels. It is often difficult to reconcile the preference of individuals with those of the aggregate group. This is known as Arrow's impossibility theorem in social choice theory (Campbell and Kelly, 2002). At the extreme, the sacrifice of individuals may ensure the survival of the whole system. Lymphocytes are not resilient individually but ensure the resilience of the immune system. Such strategies are apparently at the opposite end of Bhutan's emphasis on individual happiness. This suggests that there may be several paths towards system resilience and/or that the level and type of resilience is also a matter of choice, given the conflicting requirements (costs versus benefits at different levels).

5.3. Fundamental values and individual strength as a basis of resilient societies

At the system level, it can be illustrated by the following examples:

- fundamental prices of assets are more stable and predictable than their bubble components, which are unstable and may lead to severe crashes;
- practical skills (farming, engineering, programming, the development of the real economy, and so on) should be better rewarded both economically and in our cultures; stakeholders should pay attention to the added-value of supporting services (financing, marketing, management, and so on) and not hesitate to shrinking and redirecting efforts when these supporting services become tyrants rather than servants of the real economy;
- hard work, persistence, tenacity and dedication should be emphasized (which is at the opposite of the common modern emphasis on the role of chance and luck, the belief in easy profits, the "American dream" now fueled by a perpetual expanding credit engine).

The implementation of the recipes for resilience designed at the system level may not all apply directly to the individual, due to differences in the goals as well as psychological and physical aspects. The rest of this subsection is focused on recipes for personal resilience and top performance, which are easy to implement by everyone. To change the world, one should start with oneself.

Section 3.3 documented that many natural systems evolved to function "at the edge of chaos", characterized by a sharp balance between the level of risks they face and costly resilience build-up. Management of social-economic systems is also striving to achieve a balance between costs of increased resilience and its benefits. But would "at the edge of chaos" be a desirable state for a human? To stay a long time close to criticality, in a kind of alarmed position, requires constant attention, give rise to worries and triggers anxiety. In the end, there is the possibility that such a critical state does not lead to an efficient allocation of resources of the body and mind, but becomes stress itself.

One should consider an additional dimension, an often neglected benefit that comes from higher resilience: resilient people are more "happy" and vice-versa. Indeed, people who feel on top of their life and who can face stress are more relaxed, enjoy more the present and live longer. More resilience promotes a more positive attitude to one's own life and to others. In contrast, those of us who are in a continuous race to face the constraints of personal and professional life live in a state of anxiety, a condition that has been accelerating in severity in recent decades as witnessed by the exploding sales of antidepressants. Research in psychology and psychiatry confirms the existence of a strong interdependence between resilience and happiness, with positive feedback loops in which higher positive mind set promotes resilience and vice-versa (Jackson and Watkin, 2004; Srivastava and Sinha, 2005; Cohn et al., 2009). In particular, positive emotions help people build lasting resources (Cohn et al., 2009). And it is how we respond to stress and hard time that determine our successes or failures, rather than the nature of the stresses themselves. This supports again the need for generic and robust recipes for building up resilience and... happiness at the individual level. In a review covering a large body of research investigations on individual resilience, Coutu (2002) extracted the three main characteristics that are most often associated with resilient people:

i a staunch acceptance of reality,
ii a deep belief that life is meaningful, and
iii an uncanny ability to improvise.

Our own experience and reflection suggest to add

iv the ability to keep an inquiring mind that questions assumptions and the status quo and
iv a strong belief that our project and endeavors will succeed.

The seven factors of resilience reviewed by Jackson and Watkin (2004) from the psychological point of view overlap with the two first items, that are the need of developing a realistic view of reality and finding meanings (or causality). Indeed, they cite the following seven factors: (a) emotion regulation, (b) impulse control, (c) causal analysis, (d) self-efficacy, (e) realistic optimism, (f) empathy and (g) reaching out. These are descriptors or traits of resilient individuals. In order to be genuinely useful however, the next step is to identify whether and how it is possible to acquire, nurture and augment these traits. We are here entering the controversial domains of psychological programs and even psychiatric treatments. We take a simple "mechanistic" approach based on the premise that the above traits do not reside in a vacuum but rather are properties of bodies and minds that can be trained. Take the example of will power. In a study of one million people quoted by Baumeister and Tierney (2012), most said that self-control was their biggest weakness. So can people build up their willpower? Or are some people just born that way? In their recent book, Baumeister (who directs the social psychology program at Florida State University) and Tierney (2012) argue that willpower is like a muscle, and like all muscles, can be exhausted through overuse, but also trained to be made stronger. We could say that a strong willpower gives benefits by a slow accumulation of small gains that grow over time. The build-up of willpower operated via a positive feedback process: the more you have, the more you use "rituals" and checklist type approaches, the better the performance, the stronger is gratification for the efforts spent, the larger the willpower, the more this continues in a virtuous loop of self-reinforcement. Baumeister and Tierney also emphasize that everything is linked together and that one energy resource is used for all kinds of acts for self-control. One could then argue that, by training and augmenting the energy source, the stronger and more energetic the body and the mind, the easier it is to develop the factors promoting resilience. In this strategy, resilience has its underpinning in the strength as well as cohesion between constitutive elements found at the level of metabolism. In a recent contribution, one of us (Sornette, 2011) has laid out seven governing principles for personal resilience and performance that we repeat for completeness. We refer to the original essay and its detailed documentation and argumentation. The seven guiding recipes for individual resilience and performance are anchored in processes that control our biological and psychological well-being. Implementing these principles require willpower, which can be augmented both by the fact of being used, as in the muscle analogy of Baumeister and Tierney (2012), and by promoting the access to more energy as the source for action.

1. *Sleep*: Rest with quality sleep for a minimum of 7-8 hours per night;
2. *Love and sex*: Cultivate the romance and relationship with your special partner; interrupt your work when needed with one minute of intense focus on the loved one, perhaps using romantic pictures of him/her to trigger happiness hormones that boosts brain performance and well-being;
3. *Deep breathing and daily exercises*: Start each of your day (no exception) with 5-10 minutes of exercises, including deep breathing-stretching followed by abdominal and finishing with a very short intense workout; perform a few 2-3 minutes of intense workouts and deep breathing at different times of your day in your office or wherever you happen to be in order to oxygen your body and refresh your brain;
4. *Water and chewing*: Drink at least 2 liters of water per day (no canned juice, no coke, no beer, no sugar) outside meals and drink minimally or not at all during meals (a small glass of red wine or cup of hot green tea is fine); "drink your food" and "eat your drinks";
5. *Fruits, unrefined products, food combination, vitamin D and sun exposure and no meat and no dairy*: Eat as much fruits with water as possible on an empty stomach during the day, avoid meat and consume only unrefined products and cereals; avoid bad food combination to avoid conflicts between alkaline versus acid foods;
6. *Power foods*: onion, garlic, lemon, kiwis, almonds, nuts, dry fruits for super-performance in time of intense demand;
7. *Play, intrinsic motivation, positive psychology and will*: rediscover the homo ludens in yourself in things small and large so that work and life become a large playground, cultivate motivation as a self-reinforcing positive feedback virtuous circle.

6. Human limits and operational solutions

6.1. Intrinsic human limits

6.1.1. Identification of stress signals and reactions to them

The analysis of the major industrial catastrophes, such as the 1986 Challenger space shuttle disaster, the explosion of the Ariane V rocket on its maiden flight in 1996, the Deepwater Horizon BP oil spill disaster that started on 20th April 2010, the Fukushima-Daichii nuclear accident in March 2011 and so on, reveals common problems in the following areas:

1. gathering information;
2. aggregating and communicating data;
3. maintaining a state of attention.

These same issues, which have been documented as underlying causes of these dramatic events, are similarly found underlying most accidents and crises in different fields of human activity, including the financial crises that started to rock the world in 2007-2008.

Gathering evidence about informative incidents is a well-known challenging task in the practice of operational risk management. Employees often experience a conflict of interests with respect to reporting problems concerning the area of their own responsibility or those of their colleagues. This may rise, for example, from the fear of punishment, disapproval of colleagues and seniors, and increase of duties to correct revealed weaknesses. As a result, signals of stress are often lost, near misses are not recorded, forgotten or dismissed, and decisions are made on the basis of unrealistically optimistic data. Furthermore, from the failure of reporting and aggregating information that is in fact known within the organization, vulnerabilities are accumulated and lead to greater accidents.

The other side of the "information problem" lies in the difficulty of maintaining a constant state of attention or excitation. It is not enough to detect a signal of growing stress, but there should be measures taken to address the issue. Unfortunately, people get used to warning signals and false alarms, and lower their guard. Again, this applies to all the above mentioned industrial catastrophes and to many more.

The first step in dealing with these problems is for the top-management to accept the unavoidable nature of stress so that appropriate stimulating mechanisms can be developed:

1. for gathering and communicating information:
    - no punishment for self-reported occasional misses, as well as in the cases when all sufficient measures were taken to ensure a desired result (i.e. evaluating the process of decision-making, but not only

an ex-post outcome);
- confidentiality;

2. for maintaining a high attention:
    - “zero tolerance” to controllable misses;
    - the introduction of random stressors (such as sending “fake hard customers”;
    - to check the professionalism of employees);
    - a rewarding system for catching a stress signal.

6.1.2. *The “logic of failure”*

In their study on the “logic of failures” (Dörner et al., 1990; Dörner, 1997), Dörner and collaborators have found that there is indeed a logic in the origins and processes leading to failures, in the sense that (a) humans experience failure more often than success when intervening in complex systems, (b) the failures are not random, but exhibit common patterns and (c) the understanding of these patterns offer operational rules to prevent the failures. The studies performed by Dörner et al. led them to formulate general recommendations taking the exact counterpoints of the negative behaviors and habits that tend to inhabit people. Unsurprisingly, these recommendations overlap and sometimes complement the generic recipes outlined in section 5.1. In order to avoid failure and develop successful management of complex systems, one should

a continue to reflect and ask questions during the evolution of the project or system,
b act after careful analysis and be multi-faceted to ensure a rich toolbox of responses,
c strive to anticipate effects of one’s actions,
d estimate possible negative feedbacks and unintended consequences,
e not shy away from adapting policies that are not working, and
f carefully assess the real goals as opposed to be over-involvement in pet projects.

6.1.3. *The “illusion of control” syndrome*

Last but not least, one should always have in mind the “illusion of control” syndrome (Langer, 1975; Satinover and Sornette, 2007; 2011), as already mentioned in the introduction. As a corollary, individuals appear hard-wired to over-attribute success to skill, and to underestimate the role of chance, when both are in fact present. Grandin and Johnson (2005) recount experiments pitting humans against rats, in which the humans, like the rats, have not been explained the rules of the game but must infer them from the situation. In such experiments, rats often beat humans, because humans tend to over-interpret randomness and find meaning in random patterns. Normal people have an “interpreter” in their left brain that takes all the random, contradictory details of whatever they are doing or remembering at the moment, and smoothes everything in one coherent story. If there are details that do not fit, they are edited out or revised for sense making, providing a powerful mechanism for the illusion of meaning and of control. These phenomena are ubiquitous. Langer (1975) summarized the problem in a rather amusing way: “normal people’s high level of general intelligence makes them too smart for their own good.”

This problem is perhaps best illustrated in finance where, after a full cycle of rise and fall after which stocks are valued just where they were at the start before the fall, most investors lose money by over-reacting and thus selling close to the bottom before the rebound (Guyon, 1965). More recently, a very large body of academic works support the conclusion that most managers underperform the “buy-and-hold” strategy and that the persistence of winners is very rare (Malkiel, 2012). Nevertheless, managed funds and the demand for professional investment advice has never been stronger and is a multi-trillion dollar industry, dominating the world of pension funds, mutual funds, sovereign funds, private banking and so on. The “illusion of control” syndrome is thus a call for realizing and understanding our cognitive biases . The psychological as well as philosophical literatures have discussed many times the intrinsic limits faced by any investigator trying to determine whether and how her own cognitive processes may deform her knowledge construction of the “outside” world. This is typified at the extreme by the madman who concludes, from the deformed lenses of his perceptions, that it is the rest of the world who is mad. In the context of dynamical game theory, Satinover and Sornette (2007; 2011) have determined precisely the conditions under which the “illusion of control” syndrome occurs. In dynamical first-entry games (a subset of game theory), they found that low entropy (more informative) strategies under-perform high entropy (random) strategies. This typically occurs in situations where there is a large amount of randomness, of uncertainty as well as the presence of negative feedbacks of the decision makers’ actions onto the system.

6.2. *“Crisis flight simulator” for management of complex systems and resilience build-up*

The “illusion of control” and the “logic of failure” raise the following fundamental questions for practice. What is the value of management? How much management and control is needed? How can we falsify the value of control and of management, given that we do not have the luxury of playing history twice or multiple times? How is it possible to improve management skills when dealing with complex systems? Many studies and thinkers have pondered these issues. The recommendations given in the literature argue for a balance between extremes, such as strong top-down leadership to convey the goals and the vision, together with large responsibility and autonomy given to the bottom execution; a cohesive and strong backbone linking the individuals in an efficient hierarchical network of complementary abilities and trust together with a flexible adaptive organization to face changing and uncertain conditions. But how to achieve the right balance?

We propose that the answer lies in fostering a permeating and ubiquitous learning and testing environment, as occurs during academic curricula, and which should grow within all resilient organizations. This can take the shape of the systematic development of "crisis flight simulators" everywhere.

Consider the subprime crisis that started in 2007 with epicenter in the U.S. and the on-going sovereign crisis in Europe. To stop these systemic crises, central banks and governments have resorted to extraordinary measures, such as growing the balance of central banks with amounts of so-called toxic assets at levels dwarfing all known historical precedents. It is fair to state that we now live in a world where central banks and government are performing experiments in real time that are impacting billions of people, based on dated economic models (such as the Dynamical Stochastic General Equilibrium), which until recently did not even incorporate a banking sector and could not consider the possibility of systemic financial failures due to contagion. Not much has changed, though. The "primitive" approach of policy and decision making, based on rule-of-thumbs, political agenda, demagogy, and untested models, is still in full force. In contrast, we argue that progress requires to endow decision makers with tools to learn and to practice at the level that airline pilots or surgeons already experience in their training. These "flight" or "surgery" simulators reproduce as faithfully as possible real processes as well as all imaginable and even unimaginable scenarios to perform "what if" exercises. This approach is relevant for all kind of decision makers, including those in the financial, policy, engineering and environmental domains, and concerns also the public, students and anyone interested and responsible. A good example of an early development of "crisis flight simulators" is the approach of Dörner et al. (1990) and Dörner (1997) mentioned above. Dörner and his colleagues conducted experiments with computer simulated environments, which included two groups of participants - executives and students. Analyzing the results of the experiments and the significant better performance of the executives, the authors proposed the concept of "strategic flexibility", which is essential in coping with uncertainty and can be learnt through practical experience or by successive computer simulations.

The goal should thus consist in developing sophisticated convivial simulation platforms that incorporate detailed physical, geological, meteorological, geological, architectural, sociological, cultural, psychological and economic data with all known (and to be tested) feedback loops. For a given simulation, decision makers are given the power to make decisions on allocated resources to develop projects and to mitigate risks according to different strategies. The simulations will then demonstrate the consequences of the decisions within a multi-period setup. Only by "living" through scenarios and experiencing them, can decision makers make progress. For instance, there is enormous evidence in the laboratory and in real life settings that veterans who have lived through financial bubbles and crashes, through environmental crises and so on, are much better at prevention and mitigation. But, in practice, the cost is too large to learn from real life crises. This calls for a methodology for resilience based on the development of simulators that decision makers use to understand the complex dynamics of out-of-equilibrium systems whose behavior intrinsically includes changes of regimes, bifurcations, tipping points and their associated crises. This ambition is for instance shared by the FuturICT project, as embodied in its "Living Earth Simulator", which aims at enabling the exploration of future scenarios by large-scale simulations and hybrid modeling approaches running on supercomputers (Bishop et al., 2011; Helbing and Balietti, 2011; Helbing et al., 2011).

With such tools, the decision maker is able to understand holistically the dynamics of the system, in a systemic way, which means that he can understand the existence of systemic instabilities as one of the dynamical solutions of the system evolution. This must be complemented by a classification of the different regimes possible, a phase diagram in which the decision maker understands which control leads to the region of the unwanted regimes and which do not. He needs to understand that bifurcations and changes of regime are a natural and expected part of natural and social systems. This understanding does not occur via studying arcane mathematical theory but, instead, by experimenting as in real life, albeit with the protective comfort of the simulator and the efficiency of scaling space and time as needed. Only under this systemic structural understanding, can he interpret correctly the precursory signs in real life and use them to correct and steer the system towards resilience and sustainability.

In order to achieve effective "crisis flight simulator" platform for management and resilience, three technical goals must be achieved: (i) modeling, (ii) collective action and (iii) crowd sourcing. First, there is the need to transform complex risks scenarios from natural language into a logical, machine-interpretable description. For that, it is necessary to reach a sufficient level of abstraction to address a broad variety of scenarios and make them reusable. We envision that complex risk scenarios could be seen as electronic circuits with components acting as relays, delayers, amplifiers, dampers, transistors, and so on, connecting at-risk entities. For instance, consider three entities A, B and C. A transistor dependence would be: A fails implies that C fails if B is activated. By combining basic components, arbitrarily complicated scenarios can be built and, moreover, scenarios can be machine-tested. This first approach intends to identify elementary components from which any arbitrarily complicated risk situation can be designed and tested in real risk situations. After preliminary calibration, volunteers can be invited to play, to reuse these elements, to build and to simulate their own risk scenarios. Second, there is the need to develop a sustainable mobilization of the crowds, so as to promote a "collective action" approach to large and systemic risks (T. Maillart, private discussions). While the first proposed approach to complex risks management might interest risk researchers and professionals, its democratized adoption by users of very different backgrounds, socioeconomic horizons, age classes and cultures is critical to gather and to organize scattered information, in order to address large scale scenarios. To ensure sustainable mobilization of large populations of users, focus on intrinsic motivation is key. It will be necessary to explore the factors of motivation (hedonic pleasure and personal interests) and their relative proportion from their contribution behaviors. Two kinds of behaviors are expected: in their personal sphere of interests, many individuals will gather and submit the necessary information to document and verify scenarios, while others will rather focus on technical challenges for the pleasure of making a nice design that works. Progressive migration from the first to the second category becomes a proxy of internalization of knowledge and skills by users. Intrinsic motivation ought to drive also individual efforts towards most relevant risk scenarios. As a consequence, having a large number of contributors is the assurance of more accurate design, of better testing and of increased validity. By having many people contribute similar scenarios (or pieces of scenarios), it will be possible to derive quantitative metrics out of qualitative contributions. Third, it is necessary to develop crowd sourcing to improve the perception of regime shifts and systemic crises. There is always a large part of subjectivity in the way people perceive risks, which are complex, uncertain or even ambiguous. Such biases are likely to emerge as more individuals with various backgrounds and interests will join and contribute to the simulation platform, and therefore, must be considered. In fact, the possibility to capture human perception biases regarding risks at large scales should rather be considered as an opportunity to understand the revealed preferences that, by self-fulfilling prophecies or reflexivity, condition the choices of society. Crowd sourcing is expected to reveal and address idiosyncratic perception biases and further extract systematic ones among large populations. Finally, with contributors coming from various cultural backgrounds, differences in the perception of risks should be empirically measured at large scales.

The simulation tools of the “crisis flight simulator” for resilience build-up should be extraordinarily useful for

i scientific synthesis of different fields in a coherent framework,
ii the training of decision makers who do not realize the unintended consequences of their decisions (many of whom are negative and often with enormously bad consequences) and
iii the education of the public, of citizens and of students to be informed as well as to help them direct policy by voting in an informed way.

Different institutions and companies have developed initiatives that have some relationship but are in general much more limited than the presently proposed vision of “crisis flight simulators”. One can mention the Japanese Earth Simulator , the Sentient World White Paper , Google.org that utilizes “collective action”, Gapminder for monitoring and visualizing various indices and others.

6.3. Resilience by multi-variable measurement and prediction

6.3.1. Multi-variable measurement of resilience

In Section 3, it was demonstrated that resilience can be seen as one of the indirect measures of stress used in social sciences. Considering a problem from a different angle, the resilience of a system, i.e., its ability to cope with stress, and its measurement can be improved by taking into account:

1. the multidimensionality of resilience, as the development of a system can be motivated by several goals (subsection 5.2);
2. complementary (preferably direct) dynamical measures of:
    - stressors, to which the system is sensitive (e.g. risk measures are used in a probabilistic approach),
    - stress, developing within the system (e.g. crash hazard rate),
    - costs and efficiency of managerial actions.

As a system is subjected to the influence of numerous factors, which have different effects and are interconnected, it is important that the measurement of resilience would be based not on a single characteristic but include an ensemble of them. It would be very useful to track the dynamics of different stressors and their influence on the stress reaction of the system, as well as monitor how managerial actions affect both of them. Armed with this type of quantitative data, decision makers will be able to better understand the regime in which the system is functioning. They will be able to identify the true source of change in the stress level of the system. The origin of change may include some beneficial dynamics of a stressor, managerial actions, and/or the adaptation of the system to changing conditions. Decision makers may then be able to develop better policy, based on a risk-benefit analysis.

Despite existing limitations, especially in systems that include the “human factor” (see subsection 6.1), theoretical and empirical findings suggest that such a complex quantitative approach to resilience is not only possible but, in many cases, can be enhanced by the development of a predicting capacity.

The next subsection 6.3.2 proposes a more systematic classification of the type of stressors. Then, subsection 6.3.3 builds on the endogenous nature of many crises to suggest the most ambitious approach yet discussed here, namely the “time@risk” approach based on the monitoring of precursors towards the prediction of financial and economic crises. This is nothing but the operational implementation of the famous maxim “Gouverner, c’est prévoir” (governing is predicting) by Emile de Girardin.

6.3.2. Analysis of stressor types (exogenous versus endogenous and their interplay)

1. Stressors can come from external sources and the environment, beyond the direct control of the system. Some are knowable, quantifiable, in the possible losses and their frequencies. This is the favorable situations where counter measures can be build to prepare for the possible losses and to catalyze recovery, using the dynamical framework described in sections 3.2 and 3.3. Considering external stressors, responsible managers and decision makers should also consider the real surprises, such as in the Knightian uncertainty of unknown unknowns popularized by Taleb (2007)'s "black swans". Then, resilience can only be attained with the interplay between, as already said, (i) individual strength and adaptation, (ii) cohesion of the social group as well as (iii) a balance between a clear top-down vision that does not exclude the empowerment of individuals at the "bottom" to be able to inform the top and act decisively when needed.
2. Stressors are also often of an endogenous nature, even if exogenous influences and fluctuating perturbations are always present in out-of-equilibrium open "living" systems. By endogenous, we mean that there is a progressive evolution and maturation of internal interactions between constitutive elements that may give rise to surprising large-scale collective changes. Mathematically, the theory of bifurcations describes well the sudden change of regime from one state or attractor to another one or to a set of other competing attractors upon the small variation of a so-called control parameter. In the bifurcation theory applied to dynamical systems, the fundamental reduction theorem states that bifurcations between states can only occur through a limited number of ways that are known and classified (Thom, 1989; Guckenheimer and Holmes, 1983; Manoel and Stewart, 2000; Kuznetsov, 2004) and under the change of a small number of (most likely, one) control parameters. Of course, what is the control parameter relevant for a given transition is not known in general but the knowledge that this is the case empowers the decision maker to realize that a given crisis may have a "simple" set of mechanisms after all, whose understanding may be used to track the transition. More precisely, according to this view, it is possible to develop advanced diagnostics of an incoming crisis and invest in techniques to identify precursors. As a corollary, resilience involves precautionary actions that address the observed internal changes. More ambitiously, managers should consider the possibility to change the course and steer the system away from the trouble that is progressively announced by the precursors. In this vein, we claim that many, if not most catastrophes, occur as a surprise because stakeholders and managers have ignored either by lack of knowledge, insufficient commitment or on purpose, the telling signs of the incoming crisis.

6.3.3. Resilience by advanced diagnostics and precautionary actions in finance and economics: the "time@risk" approach

Imagine you had advanced warning signs (and that you listened to them) about the future occurrence of an adverse shock to your firm. Imagine that you could have access to precursory signs of diseases not yet symptomatic in your body (as is the dream of Proteomics). Imagine you could rely on an indicator diagnosing the existence of a financial bubble and indicating the probable time of its burst . Imagine that these advanced signs would be revealed years in advance. With this kind of information, you could prepare, you could reflect on what is not working and what could be improved or changed. You could start a process towards building stronger resilience, catalyzed by the knowledge of the nature and severity of the stressors forecasted to come. In contrast to ignorance or complacency, advanced diagnostics could revolutionize risk management by pushing us into action to build defenses. A working advanced diagnostic system would not be static, but would provide continuous updates on possible scenarios and their probabilistic weights, so that a culture of preparedness and adaptation be promoted. This corresponds to exploiting the concept elaborated in section 4 concerning the coevolution of systems and their stressors. Here, we go one step further by suggesting that forecasting the occurrence of crises promotes the evolution of the system towards a higher level of resilience that could not be achieved even by evolution (which is backward looking). Advanced diagnostics of crises constitutes the next level of evolution for cognizant creatures who use advanced scientific tools to forecast their future.

To be concrete, we describe how this system, which we refer to as the "time@risk" approach, would look like when targeting financial and economic instabilities. Here, the outstanding challenge is to develop predictions of systemic risk and global financial instabilities that have emerged as leading concerns in modern economies and with globalization. As Einstein said: "Problems cannot be solved by the same level of thinking that created them." Therefore, a truly interdisciplinary approach to the diagnostic of such crises is required. By leveraging on expertise in Economics, Mathematics, Statistical Physics and Computer Science, a novel integrated and network-oriented approach can be brought to bear on the issue. This would require providing

1. a theoretical framework to measure systemic risk in global financial market and financial networks;
2. an ICT collaborative platform for monitoring global systemic risk;
3. algorithms and models to forecast and visualize interactively possible future scenarios.

Consider the example of a financial crash, such as "Black Monday" 19 October 1987 mentioned in section 4.4.1. A sum of evidences suggests that it did not come out the blue. Postmortem analysis of many financial crashes shows the development of a kind of standard scenario, as documented for instance by Kindleberger (2005) and Sornette (2003). A financial crash is the result of increasing financial leverage developing together with social herding and the psychology of a "new economy". Specifically, this creates bubbles, and the crashes are nothing but the termination and burst of the bubbles. Using the concept of stress developed throughout the present essay, this endogenous maturation of the financial system towards an instability can be quantified by the excess super-exponential accelerating bubble price. This excess growing price can be used as a *direct measure of the level of stress* increasing within the system. This can be shown via the theoretical linkage between the "crash hazard rate" and the excess price (Johansen et al., 1999; 2000; Yan et al., 2012).

Other *early warning stress signals and diagnostics* for the upcoming transition into the major regime shifts associated with crises include, as reported by Sornette (2002; 2004), Dakos et al. (2008), Scheffer (2009),

- i a slowing down of the recovery from perturbations,
- ii increasing or decreasing autocorrelations,
- iii increasing variance of endogenous fluctuations,
- iv appearance of flickering and stochastic resonance, and other noise amplification effects (Harras et al., 2012),
- v increasing spatial coherence, and singular behavior of metrics revealing positive feedbacks (Sammis and Sornette, 2002; Johansen and Sornette, 2010).

This is a very general problem and, in principle, the "time@risk" approach can be extended to various domains of application. The corresponding "time@risk" platform should ideally

- a signal the possible occurrence of a crisis;
- b provide insights to adopt the appropriate policy measures; and
- c allow evaluating future scenarios according to the chosen policy.

The development of a framework for a computational forecasting infrastructure must necessarily combine modeling the relevant entangled networks with empirical analysis and validation of the models. Finally, there is a need to craft the tools into an interactive platform. Therefore, the objectives of the "time@risk" approach can be stated as follows.

1. Provide novel indicators and methods to estimate the origin and dynamics of systemic risk and forecast probability of systemic crises.
2. Develop agent-based models of the interacting networks which (a) are suitable to be validated, and (b) allow to compute indicators of systemic risk.
3. Validate the models with empirical data.
4. Develop a measurement platform in which it is possible to
    - a load and share relevant data about the involved institutions and their relations,
    - b produce topical maps of interacting networks,
    - c detect the propagation of distress, and
    - d perform simulations, scenario analysis, and systemic risk estimation.

This is an ambitious and risky approach. One should be aware of the risks and difficulties in the development of such a computational forecasting framework. For this reason, tasks should be developed both at empirical and modeling levels and with resources including a collaborative team of experts in an interdisciplinary atmosphere, forecasting technologies combined with the science of networks in order to validate the results obtained. In this way, the following insights can be implemented.

1. In contrast with a majority view of the current understanding, the global industrial, economic, financial and ecological systems are complex in which (a) micro and macro behavior can be dramatically different, (b) density and heterogeneity of the links as well as the whole topology (clusters, cycles and other patterns) may play a role on the (in)stability of the system and (c) time evolution is crucial for spillover effects and externalities to cascade across the system. In this context, equilibrium approaches deliver useful but insufficient and sometimes fundamentally misleading and dangerous insights.
2. It is useful to develop an integrated micro-macro approach including an analysis of a mesoscopic scale in which the system under study is seen as a network of different sectors (e.g. business lines such as commercial banks, investment banks, mutual funds, insurance companies, etc.) with a varying degree of interdependence among them.
3. One can leverage the deep knowledge recently gained by the complex networks community about failure cascades (Buldyrev et al., 2010) and contagion in networks.
4. It is necessary to go beyond the idea, dominant for long times, that big crises need big shocks and offer quantitative understanding of endogenous mechanisms of onset and amplification of crises. In this view, systemic risk is fundamentally different and possibly at odds with individual risk (e.g Morris and Shin 2008, Brunnermeier 2009). In particular, local shocks can also have systemic repercussions (Delli Gatti et al. 2005, Iori et al. 2008; Battiston et al 2007; Sieczka, Sornette and Holyst, 2011).

In the economic and financial applications, the list can be enhanced by the following objectives.

5. A necessary goal is to challenge the mainstream economics vision that more links (and thus interdependence) make always the economy more stable (Allen and Gale, 2000; Shiller, 2004; 2008; Merton and Bodie, 2005). Unfortunately, under some not so infrequent

circumstances, financial integration may increase systemic risk (Lorenz and Battiston 2008; Battiston et al 2009). More generally, it has been shown that stronger coupling leads to increased risks of synchronization and to the occurrence of system-wide catastrophes (Sornette, 1994; Osorio et al., 2010). Such events have been termed "dragon-kings" to emphasize their special impact and the specific generating mechanisms (Sornette, 2009; Sornette and Ouillon, 2012).

6. A promising approach is to combine Minsky (1982)'s view, currently under re- evaluation, of an endogenous build-up of financial fragility in the economy with a network approach. As a result, the extent of the systemic repercussions at the Minsky moment depends not only on the distribution of fragility across the agents but also on the structure of their network of mutual financial exposures.
7. It is important to complement the panorama of projects trying to identify precursors of crises from stock prices dynamics, by focusing instead on the network of exposures among financial institutions, which play a crucial role in the spreading out of financial distress, both in the Money Market (e.g., interbank, Repo, and so on, with maturity < 1 year), in the Capital Market (e.g., bonds, long-term loans, etc. > 1 year) and possibly in the OTC derivatives market.

7. Concluding remarks

Ideally, an individual, a group or a society would like to be optimized fully for the present, enjoying now the comfort resulting from past achievements and investments while, at the same time, be prepared for the inevitable future stressors that are difficult to foresee. The concept of resilience embodies the quest towards the ability to sustain shocks, be they externally or internally generated or both, to suffer from these shocks as little as possible, for the shortest time possible, and to recover with the full functionalities that existed before the perturbation. Building up resilience is, like risk management, confronted with the eternal conflict between the long-term benefits and the short-term costs. Indeed, building up resilience is costly, as it swallows resources that would otherwise be directed towards optimal present output. And like in risk management, the benefits are visible only when a serious crisis hits the system, which sometimes occur only over time scales of decades. The level of efforts towards resilience can thus be seen to be fundamentally anchored in a kind of philosophical perspective of one's personal life for the individual, or a choice of culture or of society for the larger group. Building up resilience can ultimately be seen as a problem of decision making in the face of conflicting evidence and goals as well as limited strengths in the presence of a complex stochastic environment, with all its complexity and entanglement with all other aspects of life and society. It is a balance between the present versus the future, between commitments for costly investments versus present enjoyments. Yukalov and Sornette (2012) have recently shown that self-organization in complex systems can be treated as decision making (as it is performed by humans) and, vice versa, decision making is nothing but a kind of self-organization in the decision maker nervous systems. Framing the build-up of resilience as a dynamical and continuous decision making process offers novel perspectives, which beg to be explored, based on the bridge between complex pattern formation and evolutionary emergence of novel properties.

References

Abbott, D. (2010): Asymmetry and disorder: a decade of Parrondo's paradox, in: Fluctuation and Noise Letters, 9 (1), 129-156.

Akaev, A., V. Sadovnichy, and A. Korotayev (2012): On the dynamics of the world demographic transition and financial-economic crises forecasts, in: Eur. Phys. J. Special Topics, 205, 355-373.

Aldwin, C.M. (2007): Stress, coping and development: an integrative perspective, The Guilford Press.

Allen, F., S. Morris, and H. S. Shin (2003): Beauty Contests, Bubbles and Iterated Expectations in Asset Markets, Cowles Foundation Discussion Paper 1406.

Archibald, J.D. et al. (2010): Cretaceous Extinctions: Multiple Causes, in: Science, 328, 973.

Aubin, J.P. (1991): Viability Theory, Birkhauser, Basel.

Axelrod, R. and W. D. Hamilton (1981): The Evolution of Cooperation, in: Science, 211 (4489), 1390-1396.

Bak, P. (1996): How nature works: the science of self-organized criticality, New York: Copernicus.

Bak, P. and M. Paczuski (1995): Complexity, contingency, and criticality, Proceedings of the National Academy of Sciences of the USA, 92 (15), 6689-6696.

Bak, P. and K. Sneppen (1993): Punctuated equilibrium and criticality in a simple model of evolution, in: Physical Review Letters, 71 (24), 4083-4086.

Barro, R. J., Fama, E. F., Fischel, D. R., Meltzer, A. H., Roll, R., and Telser, L. G. (1989): Black Monday and the Future of Financial Markets, in: R. W. Kamphuis, Jr., R. C. Kormendi, and J. W. H. Watson (ed.), Mid American Institute for Public Policy Research, Inc. and Dow Jones-Irwin, Inc. Toronto, Canada.

Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. C. N. and Stiglitz, J. E. (2007): Credit Chains and Bankruptcies Avalanches in Production Networks, in: Journal of Economic Dynamics and Control, 31, 2061-2084.

Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. C. N. and Stiglitz, J. E. (2007): Credit Chains and Bankruptcies Avalanches in Production Networks, in: Journal of Economic Dynamics and Control, 31, 2061-2084.

Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. C. N. and Stiglitz, J. E. (2009): Liaisons Dangereuses: Increasing Connectivity, Risk Sharing and Systemic Risk, in: Journal of Economic Dynamics and Control, (2012) doi:10.1016/j.jedc.2012.04.001, Early version: NBER Working Paper No. w15611.

Baumeister, R.F. (2010): Is there anything good about men?: how cultures flourish by exploiting men, Oxford University Press, New York.

Baumeister, R.F. and J. Tierney (2012): Willpower: Rediscovering the Greatest Human Strength, Penguin Books.

Bertrand, K.Z., G. Lindsay, Y. Bar-Yam (2012): Food briefing, NECSI Report 2012-08-01 (8/15/2012).

Birkmann, J. (2006): Measuring Vulnerability to Natural Hazards: Towards Disaster Resilient Societies, United Nations University Press, Tokyo.

Bishop, S., D. Helbing, P. Lukowicz, and R. Conte (2011): FuturICT: FET Flagship Pilot Project, in: Procedia Computer Science, 7, 34-38.

Bollinger, R.R., A.S. Barbas, E.L. Bush, S.S. Lin and W. Parker (2007): Biofilms in the large bowel suggest an apparent function of the human vermiform appendix, in: J Theor Biol in press, doi:10.1016/j.jtbi.2007.08.032.

Bowles, S. and H. Gintis (2004): The evolution of strong reciprocity: cooperation in heterogeneous populations, in: Theoretical Population Biology, 65 (1), 17-28.

Boyd, J.R. (1986): Patterns of conflicts, Mimeo, 196 pp.

Bruneau, M., S.E. Chang, R.T. Eguchi, G.C. Lee, T.D. O'Bourke, A.M. Reinhorn, M. Shinozuka, K. Tierney, W.A. Wallace and D. von Winterfeld (2003): A framework to quantitatively assess and enhance the seismic resilience of communities, in: Earthquake Spectra, 19, 733-752.

Brunnermeier, M.K. (2009): Deciphering the Liquidity and Credit Crunch 2007-2008, Journal of Economic Perspective 23 (1), 77-100.

Buldyrev, S.V., R. Parshani, G. Paul, H.E. Stanley and S. Havlin (2010): Catastrophic cascade of failures in interdependent networks, in: Nature, 464, 1025-1028.

Campbell, D.E. and Kelly, J.S. (2002): Impossibility theorems in the Arrovian framework, in Handbook of social choice and welfare (ed. by Kenneth J. Arrow, Amartya K. Sen and Kotaro Suzumura), volume 1, pages 35–94, Elsevier.

Campbell, W.J. (2011): 'Too early to say': Zhou was speaking about 1968, not 1789, Media Myth Alert, (`http://mediamythalert.wordpress.com/2011/06/14/too-early-to-say-zhou-was-speaking-about-1968-not-1789/`), accessed 14 June 2011.

Cederman, L.-E., C. Warren and D. Sornette (2011): Testing Clausewitz: Nationalism, Mass Mobilization, and the Severity of War, in: International Organization, 65 (4), 605-638.

Challet, D., S. Solomon, and G. Yaari (2009): The Universal Shape of Economic Recession and Recovery after a Shock, in: Economics: The Open-Access, Open-Assessment E-Journal, 3, 2009-2036.

Chang, S.E. and M. Shinozuka (2004): Measuring improvements in the disaster resilience of communities, in: Earthquake Spectra, 20 (3), 739-755.

Chen, W.-F., D.-J. Han (1988): Plasticity for structural engineers, J. Ross Publishing (2007). Originally published, Springer-Verlag.

Chialvo, D. (2006): Are our senses critical? In: Nature Physics, 2, 301-302.

Ciliberti, S., O.C. Martin, and A. Wagner (2007): Innovation and robustness in complex regulatory gene networks, in: Proceedings of the National Academy of Sciences USA, 104(34), 13591-13596.

Clausewitz, Carl von (1984) [1832]: On War, Revised edition edited and translated by Michael Howard and Peter Paret, Princeton University Press, Princeton, N.J.

Cohn, M.A., B.L. Fredrickson, S.L. Brown, J.A. Mikels and A.M. Conway, Happiness unpacked: Positive emotions increase life satisfaction by building resilience, Emotion 9(3), 361-368 (2009).

Contrada, R. and A. Baum (ed.) (2010): The Handbook of Stress Science: Biology, Psychology, and Health, in: Springer Publishing Company, 1 edition.

Cooper, C.L, P.J. Dewe, M.P. O'Driscoll (2001): Organizational Stress: A Review and Critique of Theory, Research, and Applications, Foundations for Organizational Science, A Sage Publications Series.

Courtillot, V. and J. McClinton (2002): Evolutionary Catastrophes: The Science of Mass Extinction, Cambridge University Press; Reprint edition.

Coutu, D.L. (2002): How resilience works, in: Harvard Business Review, May, 46-55.

Crane, R. and D. Sornette (2008): Robust dynamic classes revealed by measuring the response function of a social system, in: Prod. Nat. Acad. Sci. USA, 105 (41), 15649-15653.

Cutter, S.L., L. Barnes, M. Berry, C. Burton, E. Evans, E. Tate, J. Webb (2008): A place-based model for understanding community resilience to natural disasters, Global Environmental Change 18, 598-606, (`http://dx.doi.org/10.1016/j.gloenvcha.2008.07.013`), accessed 10.10.2012.

Dakos, V., M. Scheffer, E.H. Nes, V. Brovkin, V. Petoukhov and H. Held (2008): Slowing down as an early warning signal for abrupt climate change, in: Proc. Natl. Acad. Sci. USA, 38, 14308-14312.

Deffuant, G. and N. Gilbert (ed.) (2011): Viability and resilience of complex systems. Concepts, Methods and Case Studies from Ecology and Society, Kluwer Academic Publishers. Boston/Dordrecht/London.

Delli Gatti, D., Guilmi, C. D., Gaffeo, E., Giulioni, G., Gallegati, M. and Palestrini, A. (2011): A new approach to business fluctuations: heterogeneus interacting agents, scaling laws and financial fragility, in: Journal of Economic Behavior and Organization, 56, 489-512.

Diamond, J. (2002): The rise and fall of the third chimpanzee: how our animal heritage affects the way we live,Vintage, London.

Diamond, J. (2004): Collapse: How Societies Choose to Fail or Succeed, Viking Adult.

Doebeli, M., C. Hauert and T. Killingback (2004): The Evolutionary Origin of Cooperators and Defectors, in: Science, 306, 859-862.

Dörner, D. (1997): The logic of failure: ecognizing and avoiding error in complex situations, Basic Books; 1 edition.

Dörner, D., P. Nixon and S. D. Rosen (1990): The Logic of Failure [and Discussion], Philosophical Transactions of the Royal Society of London, Series B, in: Biological Sciences, 327 (1241), 463-473.

Eldredge, N. and S. J. Gould (1985): Punctuated equilibria: an alternative to phyletic gradualism, in: T.J.M.

Favre, M. and D. Sornette (2012): Strong gender differences in reproductive success variance, and the times to the most recent common ancestors, in: Journal of Theoretical Biology, 310, 43-54.

Ferguson, N. (2011): Civilization: The West and the Rest, The Penguin Press.

Gore, A. (2007): The Assault on Reason, Penguin Press HC.

Gould, S.J. (2002): The Structure of Evolutionary Theory, Belknap Press.

Gould, S.J. and N. Eldredge (1993): Punctuated equilibrium comes of age, in: Nature, 366 (6452), 223–227.

Graeber, D. (2011): Debt: The First 5,000 Years, Melville House, First Edition.

Grandin, T. and C. Johnson (2005): Animals in translation, Scribner, New York.

Guckenheimer, J. and P. Holmes (1983): Nonlinear Oscillations, Dynamical systems and Bifurcations of Vector Fields, Springer.

Gunderson, L. H. and C. S. Holling (2002): Panarchy: understanding transformations in human and natural systems, Island Press.

Gyu, Hwang Chang (2004), President, Samsung Semiconductor, BusinessWeek, 25 October 2004.

Guriev, S. and E. Zhuravskaya (2009): (Un)Happiness in Transition, in: Journal of Economic Perspectives, 23 (2), 143-168.

Guyon, D. (1965): One Way Pockets, Fraser Publishing Co.; Repr 1917 edition.

Hannan, M.T. and J. Freeman (1977): The Population Ecology of Organizations, in: American Journal of Sociology, 82 (5), 929-964.

Harmer, G.P. and D. Abbott (2002): A review of Parrondo's paradox, in: Fluctuation and Noise Letters, 2 (2), R71-R107.

Harras, G., C. Tessone and D. Sornette (2012): Noise-induced volatility of collective dynamics, in: Phys. Rev., E 85, 011150.

Helbing, D. and S. Balietti, (2011): From social data mining to forecasting socio-economic crises, in: Eur. Phys. J., Special Topics 195, 3-68.

Helbing, D., S. Balietti, S. Bishop, and P. Lukowicz (2011): Understanding, creating, and managing complex techno-socio-economic systems: Challenges and perspectives, in: Eur. Phys. J., Special Topics 195, 165-186.

Henrich, N. and J. Henrich. (2007): Why Humans Cooperate: A Cultural and Evolutionary Explanation, Evolution and Cognition series, Oxford University Press, Oxford and New York.

Hetzer, M. and D. Sornette (2011): The co-evolution of fairness preferences and altruistic punishment,PLoS ONE 8 (3), e54308, pp. 1-18 (2013)(`http://ssrn.com/abstract=1468517`), accessed 10.10.2012.

Hetzer, M. and D. Sornette (2011): An Evolutionary Model of cooperation, Fairness and Altruistic Punishment in Public Good Games, PLoS ONE (in press) (`http://ssrn.com/abstract=1927919`), accessed 10.10.2012.

Hetzer, M. and D. Sornette (2012): Coevolution of Cooperation and Altruistic Punishment by Migration and Between-Group Selection Pressure, working paper.

Hite, J.M. and W.S. Hesterly (2001): The evolution of firm networks: from emergence to early growth of the firm, in: Strategic Management Journal: 22 (3), 275-286.

Holland. J.H. (1992): Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, MIT Press, Cambridge, Massachusetts, 2nd edition, 1992. First edition, Ann Arbor, Michigan: University of Michigan Press (1975).

Hosenpud, J.D. and B.H. Greenberg (ed.) (2006): Congestive Heart Failure, Lippincott Williams & Wilkins, Third edition.

Imhof, L. A., D. Fudenberg and M.A. Nowak (2005): Evolutionary cycles of cooperation and defection, in: Proceedings of the National Academy of Sciences of the United States of America, 102 (31), 10797-10800.

Intriligator, M.D. (1997): The Role of Institutions in the Transition to a Market Economy, in: T. Haavisto, (ed.), The transition to a market economy: Transformation and reform in the baltic states, Edward Elgar Publishing.

Intriligator, M.D. (1998): Russia: lessons of the economic collapse, report to the office of the Chief Economist of the World Bank.

Iori, G., G. De Masi, O.V. Precup, G. Gabbi and G. Caldarelli (2008): A network analysis of the Italian overnight money market, in: Journal of Economic Dynamics and Control, 32 (1), 259-278 (2008).

Jackson, R. and C. Watkin (2004): The resilience inventory: Seven essential skills for overcoming life's obstacles and determining happiness, in: Selection & Development Review, 20 (6), 13-17.

Jensen, H.J. (1998): Self-Organized Criticality, Cambridge: Cambridge University Press.

Johansen, A., O. Ledoit and D. Sornette (2000): Crashes as critical points, in: International Journal of Theoretical and Applied Finance, 3 (2), 219-255, (`http://xxx.lanl.gov/abs/cond-mat/9810071`), accessed 10.10.2012.

Johansen, A. and D. Sornette (2001): Finite-time singularity in the dynamics of the world population and economic indices, in: Physica A, 294 (3-4), 465-502.

Johansen, A. and D. Sornette (2010): Shocks, Crashes and Bubbles in Financial Markets, in: Brussels Economic Review (Cahiers economiques de Bruxelles), 53 (2), 201-253 (`http://papers.ssrn.com/paper.taf?abstract_id=344980`), accessed 10.10.2012.

Johansen, A., D. Sornette and O. Ledoit (1999): Predicting Financial Crashes using discrete scale invariance, in: Journal of Risk, 1 (4), 5-32 (`http://xxx.lanl.gov/abs/cond-mat/9903321`), accessed 10.10.2012.

Kaplan, S., B.J. Garrick (1981): On the quantitative definition of risk, in: Risk Analysis, 1 (1), 11-27.

Kauffman, S. (1993), Origins of Order: Self-Organization and Selection in Evolution, Oxford University Press.

Kerr, N.L. (1983), Motivation losses in small groups: A social dilemma analysis. In: Journal of Personality and Social Psychology, 45, 819-828.

Kimura, M. (1983): The Neutral Theory of Molecular Evolution, Cambridge University Press, Cambridge.

Kindleberger, C.P. (2005): Manias, Panics, and Crashes: A History of Financial Crises, Palgrave Macmillan, 5th edition.

Kranner, I., F.V. Minibayeva, R.P. Beckett and C.E. Seal (2010): What is stress? Concepts, definitions and applications in seed science, in: New Phytologist, 188 (3), 655-673.

Kuznetsov, Yu.A. (2004): Elements of Applied Bifurcation Theory, Springer, 3rd edition.

Lagi, M., K.Z. Bertrand, Y. Bar-Yam (2011): The Food Crises and Political Instability in North Africa and the Middle East, arXiv:1108.2455.

Laibson, D. (1998), Life-cycle consumption and hyperbolic discounting, in: European Economic Review, 42, 861-871.

Langer, E.J. (1975): The Illusion of Control, Journal of Personality and Social Psychology 32 (2), 311–328.

Levina, A., J.M. Herrmann and T. Geisel (2007): Dynamical synapses causing self-organized criticality in neural networks, in: Nature Physics, 3, 857-860.

Lipsitz, L.A. (2002): Dynamics of Stability: The Physiologic Basis of Functional Health and Frailty, in: Journal of Gerontology: Biological sciences, 57A (3), B115-B125.

Lorenz, J. and Battiston, S. (2008): Systemic risk in a network fragility model analyzed with probability density evolution of persistent random walks, in: Networks and Heterogeneous Media, 3 (2), 185-200.

Lyne, J. and H. Howe (2007): Punctuated Equilibria: Rhetorical Dynamics of a Scientific Controversy, in: Harris, R. A. (ed.), Landmark Essays on Rhetoric of Science, Mahwah NJ: Hermagoras Press, p. 73.

MacKenzie, D. (2008): An Engine, Not a Camera (How Financial Models Shape Markets), The MIT Press.

Malkiel, B.G. (2012): A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing (Tenth Edition), W. W. Norton & Company; Tenth Edition edition.

Manoel, M. and I. Stewart (2000): The classification of bifurcations with hidden symmetries, in: Proc. London Math. Soc., 80 (3), 198-234.

Mauldin, J. and J. Tepper (2011): Endgame: The End of the Debt Supercycle and How It Changes Everything, Wiley.

Meisel, C., Storch, A., Hallmeyer-Elgner, S., Bullmore, E. and Gross, T. (2012): Failure of Adaptive Self-Organized Criticality during Epileptic Seizure Attacks, in: PLoS Comput Biol, 8 (1), (e1002312. doi:10.1371/journal.pcbi.1002312), accessed 18.08.2012.

Merton, R.C. and Z. Bodie (2005): Design of financial systems: towards a synthesis of function and structure, in: Journal of Investment Management, 3 (1), 1-23.

Minsky, H.P. (1982): Can "It" Happen Again?: Essays on Instability and Finance, M.E. Sharpe, editor.

Minsky, H. (2008): Stabilizing an Unstable Economy, McGraw-Hill, 1 edition.

Morris, S. and Shin, H. (2008): Financial Regulation in a System Context, in: Brookings Panel on Economic Activity, 2, 229-274.

Nechad, H., A. Helmstetter, R. El Guerjouma and D. Sornette (2005): Creep Ruptures in Heterogeneous Materials, in: Phys. Rev. Lett., 94, 045501.

Nelson, P.W. and A.S. Perelson (2002): Mathematical analysis of delay differential equation models of HIV-1 infection, in: Math Biosci 179, 73-94.

Nowak, A., M. Kus, J. Urbaniaka and T. Zarycki (2000): Simulating the coordination of individual economic decisions, in: Physica A, 287, 613-630.

Ogas, O. and S. Gaddami (2011): A billion wicked thoughts: what the world's largest experiment reveals about human desire, Dutton, Penguin Group (USA) Inc., London.

Osorio, I., M. G. Frei, D. Sornette, J. Milton and Y.-C. Lai (2010): Epileptic Seizures, quakes of the brain? In: Physical Review, E 82 (2), 021919.

Ostrom, E. (1990): Governing the Commons: The Evolution of Institutions for Collective Action, Cambridge University Press.

Pant, R. and K. Barker (2012): Building dynamic resilience estimation metrics for interdependent infrastructures, Proceedings of the European Safety and Reliability Conference, Helsinki, Finland, June 2012.

Parry M., C. Rosenzweig, A. Iglesias, G. Fischer, M. Livermore (1999): Climate change and world food security: a new assessment, in: Global Environmental Change 9, S51-S67, (`http://dx.doi.org/10.1016/S0959-3780(99)00018-7`), accessed 10.10.2012.

Perelson, A.S. (2002): Modelling viral and immune system dynamics, in: Nature, 2, 28-36.

Perelson A.S. and G. Weisbuch (1997): Immunology for physicists, in: Rev. Mod. Phys., 69, 1219-1268.

Pimm, S.L. and J.H. Lawton (1977): Number of trophic levels in ecological communities, in: Nature, 268, 329-331.

Plenz, D. (2012): Neuronal avalanches and coherence potentials, in: Eur. Phys. J., Special Topics 205, 259-301. Reinhart, C.M. and K.S. Rogoff (2011): This Time Is Different: Eight Centuries of Financial Folly, Princeton University Press.

Richards, C. (2004): Certain to Win: The Strategy of John Boyd, Applied to Business, Xlibris, Corp.

Ritsner, M.S., Brain Protection in Schizophrenia, Mood and Cognitive Disorders, Springer (May 19, 2010).

Sammis, C.G. and D. Sornette (2002): Positive Feedback, Memory and the Predictability of Earthquakes, in: Proceedings of the National Academy of Sciences USA 99 (SUPP1), 2501-2508.

Satinover, J.B. and D. Sornette (2007): "Illusion of Control" in Minority and Parrondo Games, in: Eur. Phys. J. B 60, 369-384.

Satinover, J. and D. Sornette (2011): Taming Manias: On the Origins, Inevitability, Prediction and Regulation of Bubbles and Crashes, chapter of the book "Governance and Control of Financial Systems: A Resilience Engineering Perspective," in: Gunilla Sundström and Erik Hollnagel (ed.), published by Ashgate Publishing Group in their Resilience Engineering Perspectives series.

Schaub, B., R. Lauener R and E. von Mutius (2006): The many faces of the hygiene hypothesis, in: J. Allergy Clin. Immunol, 117(5),969-977.

Scheffer, M. (2009): Critical Transitions in Nature and Society, Princeton Studies in Complexity, Princeton University Press.

Schulte, E., Z.P. Fry, E. Fast, S. Forrest and W. Weimer (2012): Software Mutational Robustness: Bridging The Gap Between Mutation Testing and Evolutionary Biology, (`http://arxiv.org/abs/1204.4224`), accessed 10.10.2012.

Schulte, P. and al. (2010): The Chicxulub Asteroid Impact and Mass Extinction at the Cretaceous-Paleogene Boundary, in: Science, 327, 1214-1218.

Schumpeter, Joseph A. (1942): Capitalism, Socialism and Democracy, London: Routledge.

Selye, H. (1973): The Evolution of the Stress Concept: The originator of the concept traces its development from the discovery in 1936 of the alarm reaction to modern therapeutic applications of syntoxic and catatoxic hormones, in: American Scientist, 61 (6), 692-699, (`http://www.jstor.org/stable/27844072`), accessed 10.10.2012.

Sheffi, Y. (2005): The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage, The MIT Press.

Shiller, R.J. (2004): The New Financial Order: Risk in the 21st Century, Princeton University Press.

Shiller, R.J. (2008): The Subprime Solution: How Today's Global Financial Crisis Happened, and What to Do about It, Princeton University Press; First Edition.

Sieczka, P., D. Sornette and J.A., Holyst (2011): The Lehman Brothers Effect and Bankruptcy Cascades, in: European Physical Journal, B 82 (3-4), 257-269.

Siegfried, T. (2005): Praise of Hard Questions, in: Science, 309 (5731), 7677.

Sornette, D. (1993): Les phénomènes critiques auto-organisés, Images de la Physique 1993, Edition du CNRS, Janvier 1994, p.9-17.

Sornette, D. (2002): Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes and human birth, in: Proceedings of the National Academy of Sciences USA 99 (SUPP1), 2522-2529.

Sornette, D. (2003): Why Stock Markets Crash (Critical Events in Complex Financial Systems), Princeton University Press.

Sornette, D. (2004): Critical Phenomena in Natural Sciences (Chaos, Fractals, Self-organization and Disorder: Concepts and Tools), 2nd ed., Springer Series in Synergetics, Heidelberg.

Sornette, D. (2005): Endogenous versus exogenous origins of crises, in monograph entitled "Extreme Events in Nature and Society", Series: The Frontiers Collection S. Albeverio, V. Jentsch and H Kantz, (ed.), Springer, Heidelberg, pp 95-119 (`http://arxiv.org/abs/physics/0412026`), accessed 10.10.2012.

Sornette, D., Dragon-Kings (2009): Black Swans and the Prediction of Crises, in: International Journal of Terraspace Science and Engineering, 2(1), 1-18, (`http://arXiv.org/abs/0907.4290` and `http://ssrn.com/abstract=1470006`), accessed 10.10.2012.

Sornette, D. (2010): Risk Management and Governance Lessons and Prospects: From The 2007-20XX Crisis, in: REVUE für Postheroisches Management, 7, 44-53.

Sornette, D. (2011): Optimization of brain and life performance: Striving for playing at the top for the long run, German version published as: "Du kannst dein Leben steigern", in: Schweizer Monat, Dezember 2011/Januar 2012, 38-49. English version published under the title "Push for Peak Performance" at the Schweizer Monat (`http://www.schweizermonat.ch/artikel/push-for-peak-performance`); eprint at `http://arxiv.org/abs/1111.4621`, accessed 18.08.2012.

Sornette, D., F. Deschatres, T. Gilbert and Y. Ageon (2004): Endogenous Versus Exogenous Shocks in Complex Networks: an Empirical Test Using Book Sale Ranking, in: Phys. Rev. Letts., 93 (22), 228701.

Sornette, D. and A. Helmstetter (2003): Endogenous Versus Exogenous Shocks in Systems with Memory, in: Physica, A 318, 577.

Sornette, D., Y. Malevergne and J.-F. Muzy (2003): What causes crashes? In: Risk, 16 (2), 67-71, (`http://arXiv.org/abs/cond-mat/0204626`), accessed 10.10.2012.

Sornette, D. and G. Ouillon (2012) editors of the special issue of Eur. Phys. J. Special Topics on "Discussion and debate: from black swans to dragon-kings - Is there life beyond power laws?", volume 25, Number 1, pp. 1-373, (`http://www.springerlink.com/content/d5x6386kw2055740/?MUD=MP`), accessed 10.10.2012.

Sornette, D. and R. Woodard (2010): Financial Bubbles, Real Estate bubbles, Derivative Bubbles, and the Financial and Economic Crisis, in: Proceedings of APFA7 (Applications of Physics in Financial Analysis), "New Approaches to the Analysis of Large-Scale Business and Economic Data," Misako Takayasu, Tsutomu Watanabe and Hideki Takayasu (ed.), Springer, (`http://arxiv.org/abs/0905.0220`), accessed 10.10.2012.

Sornette, D., R. Woodard and W.-X. Zhou (2009a): The 2006-2008 Oil Bubble: evidence of speculation, and prediction, in: Physica, A 388, 1571-1576.

Sornette, D., V.I. Yukalov, E.P. Yukalova, J.-Y. Henry, D. Schwab, and J.P. Cobb (2009b): Endogenous versus Exogenous Origins of Diseases, in: Journal of Biological Systems, 17 (2), 225-267.

Srivastava, S. and A.K. Sinha (2005): Resilience for well-being: The role of experiential learning, in: Psychological Studies, 50(1), 40-49.

Steffen, W., A. Sanderson, P. Tyson, J. Jäger, P. Matson, B. Moore III, F. Oldfield, K. Richardson, H.J. Schellnhuber, B.L. Turner II and R.J. Wasson (2004): Global change and the Earth system (a planet under pressure), Springer, Heidelberg.

Taleb, N.N. (2007): The Black Swan: The Impact of the Highly Improbable, Random House.

Taleb, N.N. (2011): Antifragility - or - The property of disorder-loving systems, Edge, (`http://www.edge.org/q2011/q11_3.html#taleb`), accessed 27.08.2012.

Taleb, N.N. and R Douady (2012): Mathematical Definition, Mapping, and Detection of (Anti)Fragility, preprint available at `http://ssrn.com/abstract=2124595`, accessed 10.10.2012.

Thom, R. (1989): Structural Stability and Morphogenesis: An Outline of a General Theory of Models, Reading, MA: Addison-Wesley.

Wagner, A. (2005): Robustness and Evolvability in Living Systems (Princeton Studies in Complexity), Princeton University Press.

Walker, B., C. S. Holling, S. R. Carpenter, and A. Kinzig (2004): Resilience, adaptability and transformability in social–ecological systems, in: Ecology and Society, 9(2), 5, (`http://www.ecologyandsociety.org/vol9/iss2/art5`), accessed 10.10.2012.

Weiten, W. and M.A. Lloyd, Psychology Applied to Modern Life: Adjustment in the 21st Century, Wadsworth Publishing; 8 edition (June 27, 2005),

Yaari, G., A. Nowak, K. Rakocy and S. Solomon (2008): Microscopic study reveals the singular origins of growth, in: Eur. Phys. J., B 62, 505-513.

Yan, W., R. Woodard and D. Sornette (2012): Inferring Fundamental Value and Crash Nonlinearity from Bubble Calibration, Quantitative Finance, DOI:10.1080/14697688.2011.606824, (e-pring at `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1719124`), accessed 10.10.2012.

Young, L.R. (1999): Artificial gravity considerations for a Mars exploration mission, in: Ann. N.Y. Acad. Sci., 871, 367-378.

Yukalov, V.I. and D. Sornette (in review): Self-organization in nature and society as decision making, in: Adv. Complex Systems(submitted).

Citation

Kovalenko, T. and Sornette, D. (2013): Dynamical Diagnosis and Solutions for Resilient Natural and Social Systems. In: Planet@Risk, 1(1): 7–33, Davos: Global Risk Forum GRF Davos.

Risk and Resilience Management in Social-Economic Systems[i]

Tatyana Kovalenko[1] and Didier Sornette[1]

[1]ETH Zurich, Department of Management, Technology and Economics

Contact: dsornette@ethz.ch

Keywords: Resilience, risk, stress, uncertainty, predictability, dynamics, exogenous, endogenous, adaptive management, black swan, dragon-king

Risk and Resilience as complementary measures of stress

We propose a definition of resilience as an important complement to risk. Both concepts describe stress within a socio-economic system from two different angles, and together allow for a comprehensive approach to governance and management. **Stress** is an internal response of a system to a perturbation called stressor or stress-factor (Kovalenko & Sornette, 2013). Here, we think of stress as a variable that characterizes the current (or potential) state of a system on a continuum scale ranging from its normal functioning state (e.g. low average level of stress with bursts below certain amplitude and time thresholds) to an unsustainable dynamics leading to a change of regime (e.g. high average stress level with strong upward trend).[ii] In natural sciences, stress can be directly quantified from its observable effects, for instance in the form of physical deformation of a stressed body in engineering or a set of common non-specific physiological changes in living biological organisms. In contrast, stress is hard to quantify in socio-economic systems. As in natural sciences, socio-economic systems are complex and multi-scaled, subjected to a large number of exogenous and endogenous factors, with feedback loops and coupling mechanisms. However, clearly differentiating responses to exogenous from responses to endogenous stressors is made harder by the existence of learning, anticipation and self-fulfilling prophecies, where beliefs govern actions with feedbacks on processes. As an alternative, an indirect approach to measure stress was developed, based on:

1) **Risk** (as the triplet of (i) probability/uncertainty, (ii) potential loss and (iii) mitigation techniques, i.e. counter-measures to reduce vulnerability of a system) characterizes possible environment- and system-specific stressors. By analogy with the Newton's third law, risk is a proxy for a potential internal stress response of a system to these threats;
2) **Resilience** (as the four-level hierarchy of (i) local 'engineering resilience', (ii) non-local 'ecological resilience', (iii) 'viability' enriched with managerial impact and (iv) adaptation and transformation mechanisms) embodies the inner capacity of a system to cope with stressors of any nature (Kovalenko & Sornette, 2013). It characterizes the maximum

---

[i] This paper is part of the IRGC Resource Guide on Resilience, available at: https://www.irgc.org/risk-governance/resilience/. Please cite like a book chapter including the following information: IRGC (2016). Resource Guide on Resilience. Lausanne: EPFL International Risk Governance Center. *v29-07-2016*

[ii] In the present paper, we do not investigate long-term effects of different levels of stress on ability of a system to respond to stressors. Prolonged extreme levels of stress may result in adverse changes of adaptive capacity of under- or overstimulated system, resembling "poverty trap" and "rigidity trap" resp. (Carpenter & Brock, 2008).

amount of stress a system can bear without a functional disruption, the system dynamics following a perturbation such as the speed of recovery of a traditional functionality, the achieved level of performance or its transformation to a completely different state.

Adding value and filling gaps with resilience

First, as their definitions deriving from their common genesis – stress – attest, resilience and risk are **closely interconnected**:

- The vulnerability of a system, being one of the constituents of risk, bridges it to resilience: indeed, the susceptibility of a system to risks and its ability to sustain stress intersect greatly and may be affected by the same managerial actions (mitigation techniques);
- When trying to balance costly universal resilience and profitable but stripping optimization, risk measures can be important indicators of a required level of resilience.

Second, resilience and risk measures are **complementary**:

- Focusing on the components of risk and resilience that can be expressed in the same units (e.g. risk exposure vs. maximum loss that a system can withstand), *comparison of their relative values* is useful to choose an appropriate response to a stressor. 'Normal' stress, when risks are significantly smaller than the system resilience, induces a 'fight' response with negative feedbacks and return to an equilibrium state. When the risk level becomes comparable to the resilience level, a 'fly' response is often initiated by employing risk-avoidance or environment-adaptation strategies. 'Extreme' stress, when resilience is insufficient, requires a major transformation of the system via positive feedback mechanisms;
- Resilience plays a distinct and crucial role in *uncertain environments* (which resonates with the IRGC view), when standard risk management techniques fail to adequately quantify or even detect existing hazards. This category includes exposure to:
    a) extreme risks, which are characterized by heavy/fat-tailed distributions with undefined mean and/or variance (e.g. existing models for operational risk are often considered to be unrealistic in capturing the peril of human failure or a cyber security breach),
    b) slow-moving risks, which are difficult to identify and monitor,
    c) surprise factors associated with Knightian uncertainty of unknown unknowns (popularized under "black swans" (Taleb, 2007));
- Finally, *complex* socio-economic systems, with nontrivial micro-macro relations, may exhibit:
    d) unsustainable dynamics and gradual maturation towards an instability leading to a bifurcation and potentially large impact events (captured under the concept of "dragon-kings" (Sornette D. , 2009), (Sornette & Ouillon, 2012)).

In any context, resilience serves as a 'safety buffer', i.e. an all-purpose resource to withstand a non-specific stress response of a system to any demand.

Instruments for resilience management

As risk and resilience are interconnected and complementary concepts, their governance and management structures may be similar, but specialized accordingly. We emphasize the following systemic elements for resilience build-up:

- clear statement of (measurable, multidimensional) *goals* to resolve conflicts of interests between time-scales (short- vs. long-term) and beneficiaries (individual vs. community);
- development - via investment, education and regulation - of *fundamental values*, right *incentives* and fair remuneration;
- strengthening of institutions for *contract enforcement*; implementation of *transparency* and *accountability* mechanisms***;***
- *diversification* and fostering of *heterogeneity*, as a reservoir of adaptive capacity;
- *decoupling* of key components to decrease systemic risk and susceptibility to cascade propagation.

Active (biological and socio-economic) systems put stress to use as a driving force of their evolution towards better fitness to changing environments. In particular, stochastic or deliberate stressors are useful for the

- identification and characterization of stress via the system response to perturbations;
- measurement of stress, e.g. via risks and resilience;
- catalysis of learning, which promotes adaptation through feedback mechanisms, and selection of specific favorable features;
- excitation of the system's readiness, maintaining an attentive and engaged state.

Depending on (i) the level of stress induced by environmental demands or endogenous processes and (ii) the degree of uncertainty/predictability of a system, we suggest four **risk and resilience management regimes**, with their corresponding **response mechanisms and management instruments (figure 1)**, which can be grouped into *two subgroups according to the stress elevation, 'normal' to 'extreme'*.

'Normal' stress, when addressed timely, usually does not endanger the very existence of a system. Negative feedbacks are appropriate and adaptation (co-evolution) of a system to (with) stressors occurs.[iii]

- **"Ad hoc management"** can be applied to cope with 'normal' stress for unpredictable complex systems in a highly uncertain environment. This regime is characterized by self-organization, decentralization of management functions and delegation of authority.
- **"Adaptive management"** (Allen & Garmestani, 2015) operates an iterative learning methodology to reduce high management uncertainty in systems with low-to-intermediate spatial and temporal variability. Within this approach, reversible repetitive interventions are preferable, which produce visible effects on a timescale of

---

[iii] As an interesting illustration, the development of advantageous attributes of human society such as cooperation and exaggerated risk taking by males have been shown to be driven by its *co-evolution* with external and internal stressors, such as competition between groups (Hetzer & Sornette, 2013), (Hetzer & Sornette, 2013) or individual males (Favre & Sornette, 2012), (Baumeister, 2010).

months to years rather than decades. Inclusiveness of stakeholders, strong leadership and community involvement enable this regime.

Extreme stressors truly determine the environmental landscape and the evolution of the system. Thus, positive feedbacks should be employed for the radical transformations needed to adapt to the new conditions.[iv] Centralization, focus on key functionality and mobilization of resources are required. The outstanding importance of extreme events is reflected in the choice of memorable names (Black swans and Dragon-kings) personifying the following regimes.

- The **"Black swan"** regime requires a management approach that deals with unpredictable exogenous disturbances of a large impact. Quantitative estimation is problematic. Critical areas should be identified and accounted for in a contingency plan; strategies to avoid most adverse trajectories must be implemented. The resilience of a system, its ability to react fast and transform when needed is essential.
- The **"Dragon-king"** regime, in contrast, suggests that certain types of extreme events are predictable. These events are the outcome of the system dynamics progressively approaching an instability leading to a transition to another mode. Monitoring and early warning signals should be a part of management practice; interventions are time-sensitive and include preparations to a possible change of course.

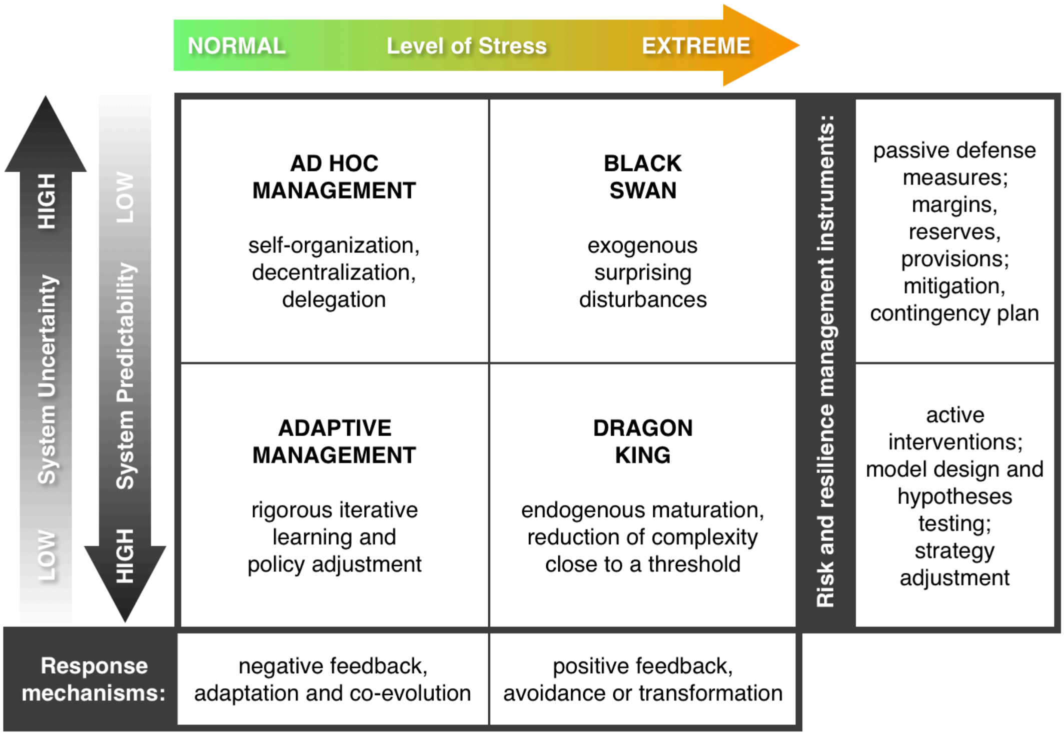

Figure 1: The four quadrants of risk and resilience management regimes corresponding to the system's degree of uncertainty/predictability and stress level within it.

[iv] For example, cardinal political and economic changes are often associated with *extreme* shocks and generic J-curve dynamics (Challet, Solomon, & Yaari, 2009), (Yaari, Nowak, Rakocy, & Solomon, 2008). This type of transitions is characterized by an initial phase of significant recession followed by a recovery, when the renewed system can outperform its preexisting level due to its better evolved fitness.

In all regimes, the resilient evolution of a socio-economic system towards a desired state requires a combination of (i) structured and strict evidence-based assessment and decision-making processes and (ii) flexibility and diversity in the considered alternative policies. The essential ingredients of management success are scientific rigor of implementation and high quality of data. (Chernov & Sornette, 2016) analyses numerous case studies and provides recommendations to facilitate knowledge acquisition and transparent communication in order to prevent distortion and the scourge of *information concealment*.

Metrics of resilience

Development of a complex system resilience calls for a multidimensional measurement approach, corresponding to multiple goals, risk factors and time scales. It includes the following steps.

1) Identification of stressors, their classification (*exo-/endo-factors*). E.g. **specific dynamical patterns observed before or after extreme events were shown to be characteristic of the (exo-/endo-) nature of the triggering factors**. This is relevant to many complex systems (Sornette & Helmstetter, 2003), (Sornette D. , 2005), and have been applied to financial shocks (Sornette, Malevergne, & Muzy, 2003), commercial sales (Sornette, Deschatres, Gilbert, & Ageon, 2004), and YouTube videos views (Crane & Sornette, 2008);
2) Quantification of **dependencies** between risk factors, with increased attention to extreme risks (Malevergne & Sornette, 2006);
3) Integration of both probabilistic measures of stress: (a) risks (observation of event probabilities, losses, vulnerability of the system) and (b) resilience ("exploration" of the stability landscape, e.g. characterized by its latitude, resistance, precariousness and panarchy (Walker, Holling, Carpenter, & Kinzig, 2004));
4) Development of direct measures of stress. E.g. for financial system, the "crash hazard rate" can be interpreted as a direct measure of the level of stress through its theoretical link to the excess bubble price (Johansen, Sornette, & Ledoit, 1999), (Johansen, Ledoit, & Sornette, 2000), (Yan, Woodard, & Sornette, 2012).
5) Quantitative measurement and characterization of the **dynamics**. E.g. different levels of resilience hierarchy can be used for a different time scales.

The following quantitative metrics pertain to each of the four risk and resilience management regimes.

- **"Ad hoc management"**. While the system is here characterized by low predictability and its stressors are stochastic, the high frequency and low severity of the latter allow for standard risk measures, such as *quantile-based approaches* (e.g. *value-at-risk* or conditional value-at-risk, i.e. *expected shortfall*), based on *historical records*, to determine adequate passive defense measures: margin levels, reserves, capital buffers, provisions, and so on.
- **"Black swan"**. The intrinsic uncertainty and the significant impact of these extreme events call for imaginative *'what-if' scenario* analysis, and prudent *stress-testing*. Option and other derivative strategies are typically put forwards for passive defense. However, these countermeasures involve risk-taking (and at the extreme gullible) counter-parties.

- **“Adaptive management”**. Carefully designed and controlled *management experiments* are iteratively maintained to determine effective, and – importantly – scalable, cost-efficient policies. The methodology emphasizes:
    - incorporation of knowledge about different aspects of the system from a broad range of *stakeholders*,
    - *model* development and formulation of alternative *testable hypotheses*,
    - carefully *monitored and controlled experimentation* to test and falsify the working hypotheses,
    - *analysis and evaluation* of the obtained data, *adjustments* of the models and management practices.
- **“Dragon-king”**. The system dynamics close to a change of regime contains early warning signals, allowing for the probabilistic estimation of the time and severity of the incoming transition. The theoretical underpinning of this predictability stems from bifurcation theory applied to dynamical systems: the fundamental reduction theorem states that, close to a change of regime, a system can transit from one state to another one only in a small number of ways, with a collapse from high to low dimensionality of the relevant variables and control parameters. These transitional “normal forms” have been systematically classified (Thom, 1989), (Guckenheimer & Holmes, 1983), (Manoel & Stewart, 2000), (Kuznetsov, 2004). The identification of the relevant control parameter(s) and the characterization of the reduced system dynamics towards a tipping point is of key importance to predict and thus prepare against extreme events in out-of-equilibrium socio-economic systems.

Annotated Bibliography

Kovalenko, T., & Sornette, D. (2013). Dynamical diagnosis and solutions for resilient natural and social systems. *Planet@Risk, 1*(1), 7–33. Retrieved from http://arxiv.org/abs/1211.1949:

- risk and resilience as complementary measures of stress
- classification of resilience measures and possible responses to stressors
- debunking “antifragility” myth
- main ingredients for the resilience of socio-economic systems

Example of a model incorporating adaptive capacity of a system as a function of its stress:

Carpenter, S. R., & Brock, W. A. (2008). Adaptive capacity and traps. *Ecology and Society, 13*(2). Retrieved from http://www.ecologyandsociety.org/vol13/iss2/art40/

Extreme events: "black swans" and "dragon-kings":

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable.* New York: Random House.

Sornette, D. (2009). Dragon-kings, black swans and the prediction of crises. *International Journal of Terraspace Science and Engineering, 2*(1), 1-18. Retrieved from http://arXiv.org/abs/0907.4290

Sornette, D., & Ouillon, G. (2012). Dragon-kings: Mechanisms, statistical methods and empirical evidence. *European Physical Journal Special Topics, 205*(1), 1-26. doi:10.1140/epjst/e2012-01559-5

Examples of co-evolution with stressors under "normal" stress and transition to a new state under "extreme" stress:

(i) cooperation:

Hetzer, M., & Sornette, D. (2013). An evolutionary model of cooperation, fairness and altruistic punishment in public good games. *PLoS ONE, 8*(11), 1-13. doi:10.1371/journal.pone.0077041

Hetzer, M., & Sornette, D. (2013). The co-evolution of fairness preferences and costly punishment. *PLoS ONE, 8*(3), 1-18. doi:10.1371/journal.pone.0054308

(ii) beneficial risk-taking of males:

Favre, M., & Sornette, D. (2012). Strong gender differences in reproductive success variance, and the times to the most recent common ancestors. *Journal of Theoretical Biology, 310*, 43-54. doi:10.1016/j.jtbi.2012.06.026

Baumeister, R. F. (2010). *Is there anything good about men?: How cultures flourish by exploiting men.* Oxford University Press.

(iii) generic J-curve dynamics:

Challet, D., Solomon, S., & Yaari, G. (2009). The universal shape of economic recession and recovery after a shock. *Economics: The Open-Access, Open-Assessment E-Journal, 3*(2009-36), 1-24. Retrieved from http://ssrn.com/abstract=1726867

Yaari, G., Nowak, A., Rakocy, K., & Solomon, S. (2008). Microscopic study reveals the singular origins of growth. *The European Physical Journal B , 62*(4), 505-513. doi:10.1140/epjb/e2008-00189-6

Allen, C. R., & Garmestani, A. S. (Eds.). (2015). *Adaptive management of social-ecological systems.* Springer Netherlands. doi:10.1007/978-94-017-9682-8:

- adaptive management framework;
- suitability criteria and implementation steps (Chapter 6 and 10);
- case studies.

Chernov, D., & Sornette, D. (2016). *Man-made catastrophes and risk information concealment: Case studies of major disasters and human fallibility.* Springer International Publishing. doi:10.1007/978-3-319-24301-6:

- 25+ case studies, including industrial, financial, social and natural catastrophes;
- 5 common factors of information concealment, viz., (i) external environment; internal environment: (ii) communication channels, (iii) risk assessment and risk knowledge management, (iv) ecology of an organization, (v) personal features of employees), and decomposing them further into 30 causes that led to the reviewed disasters.

Dynamical characterization of exogenous and endogenous factors, and its applications:

Sornette, D., & Helmstetter, A. (2003). Endogenous versus exogenous shocks in systems with memory. *Physica A, 318*(3-4), 577-591. doi:10.1016/S0378-4371(02)01371-7

Sornette, D. (2005). Endogenous versus exogenous origins of crises. (S. Albeverio, V. Jentsch, & H. Kantz, Eds.) *in the monograph "Extreme Events in Nature and Society", Series: The Frontiers Collection*, 95-119. Retrieved from http://arxiv.org/abs/physics/0412026

Sornette, D., Malevergne, Y., & Muzy, J. F. (2003). What causes crashes? *Risk, 16*(2), 67-71. Retrieved from http://arXiv.org/abs/cond-mat/0204626

Sornette, D., Deschatres, F., Gilbert, T., & Ageon, Y. (2004). Endogenous versus exogenous shocks in complex networks: An empirical test using book sale ranking. *Physical Review Letters, 93*(22), 228701. doi:10.1103/PhysRevLett.93.228701

Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Nat. Acad. Sci. USA, 105*(41), 15649-15653. doi:10.1073/pnas.0803685105

Quantification of risk factors dependences:

Malevergne, Y., & Sornette, D. (2006). *Extreme financial risks: From dependence to risk management.* Berlin Heidelberg: Springer-Verlag. doi:10.1007/b138841

Characterization of a stability landscape by its latitude, resistance, precariousness and panarchy:

Walker, B., Holling, C. S., Carpenter, S. R., & Kinzig, A. (2004). Resilience, adaptability and transformability in social–ecological systems. *Ecology and Society, 9*(5), 5. Retrieved from http://www.ecologyandsociety.org/vol9/iss2/art5/

“Crash hazard rate” as a direct measure of stress in financial systems:

Johansen, A., Ledoit, O., & Sornette, D. (2000). Crashes as critical points. *International Journal of Theoretical and Applied Finance, 3*(2), 219-255. Retrieved from http://arXiv:cond-mat/9810071v2

Johansen, A., Sornette, D., & Ledoit, O. (1999). Predicting financial crashes using discrete scale invariance. *Journal of Risk, 1*(4), 5-32. Retrieved from http://arXiv:cond-mat/9903321v3

Yan, W., Woodard, R., & Sornette, D. (2012). Inferring fundamental value and crash nonlinearity from bubble calibration. *Quantitative finance, 14*(7), 1273-1282

Bifurcation theory applied to dynamical systems: the fundamental reduction theorem and “normal forms” of transitions:

Thom, R. (1989). *Structural stability and morphogenesis: An outline of a general theory of models.* Reading, MA: Addison-Wesley.

Guckenheimer, J., & Holmes, P. (1983). *Nonlinear oscillations, dynamical systems and bifurcations of vector fields.* Springer.

Kuznetsov, Y. A. (2004). *Elements of applied bifurcation theory* (3rd ed.). Springer.

Manoel, M., & Stewart, I. (2000). The classification of bifurcations with hidden symmetries. *Proc. London Math. Soc., 80*(3), 198-234.

Discussion

The two key propositions that have been put forward in this chapter are:

- a four-level resilience hierarchy, which represents an inclusive relation: engineering resilience ⊂ ecological resilience ⊂ viability ⊂ adaptability/transformability;
- a framework - "4 quadrants" of risk severity and system control, - which identifies four regimes of risk and resilience management, the corresponding response mechanisms and instruments.

Though they were developed independently, these two theoretical constructs have a deep meaningful connection to each other. Firstly, both of them take into account underpinning stress dynamics. Different management regimes, as well as resilience types are associated with certain levels of stress, which varies from normal to extreme. Secondly, aligning management regimes with the resilience hierarchy has practical implications. The latter is instrumental for the former, i.e. this strategic mapping allows to identify the resilience approach (methods, measures, quantities, etc.) that is relevant for each regime of management.

The correspondence between the four levels of the resilience hierarchy and the "4 quadrants" of risk severity and system control is presented on figure 2.1. These two pieces, put together, complete our holistic Risk-Resilience (R-R) management system.

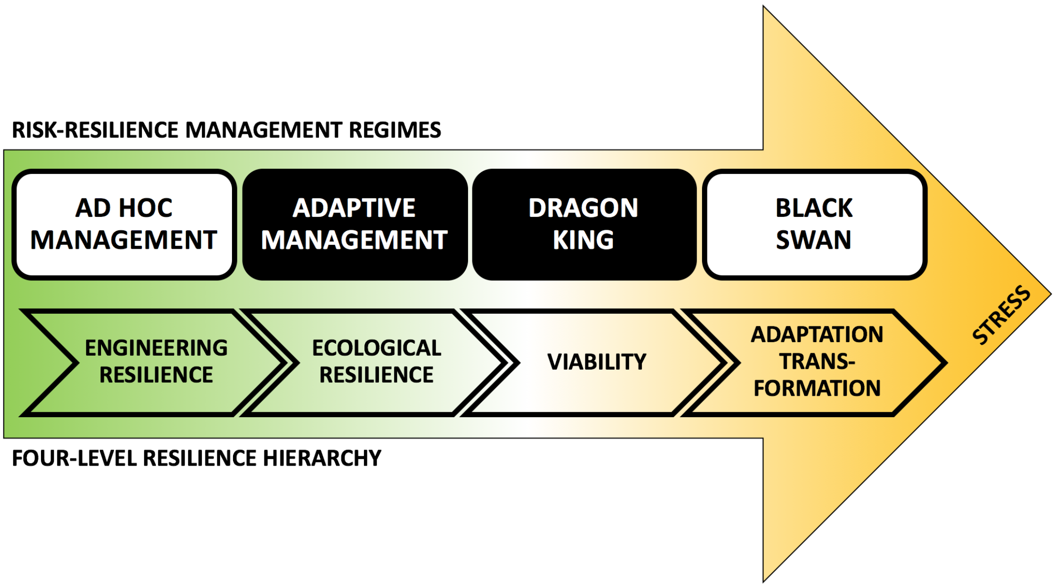

Figure 2.1: Risk-Resilience (R-R) framework: correspondence between management regimes ("4 quadrants" of risk severity and system control) and a four-level resilience hierarchy. Control levels within a management regime are indicated by color (low predictability - white, high predictability - black), and stress level increases along the background arrow from normal to extreme.

The R-R management framework unifies conceptual, methodological and diagnostic tools, which originated from a broad spectrum of scientific and business areas and previously were considered separately. The new R-R approach allows one to see them as elements of the same framework, pertaining to different regimes of a system. The management regime depends on the level of stress (severity of a stressor) and predictive/control possibility.

Our hope is that the Risk-Resilience framework can serve not only as a theoretical formulation, but also as a general management tool. At all stages of risk and resilience management (design/operation/revision), a potential/ongoing functioning regime of a system can be classified according to the "4 quadrants" framework, and the corresponding resilient response can be prepared/implemented. The R-R approach can help improving corporate analytics and reporting.

The deployment of such an ambitious R-R system in practice is very challenging. The implementation of a top-level concept may seem an opening of Pandora's box. It is especially true for the resilience system, given the diversity of localized approaches. The development of a widely-accepted standard of resilience management, similar to the existing risk management standards, could facilitate an interdisciplinary harmonization and spreading of best resilience practice. In (Häring et al., 2017) **(3)** this view is pushed forward by proposing a generic resilience management process. Figure 2.2 shows the process cycle.

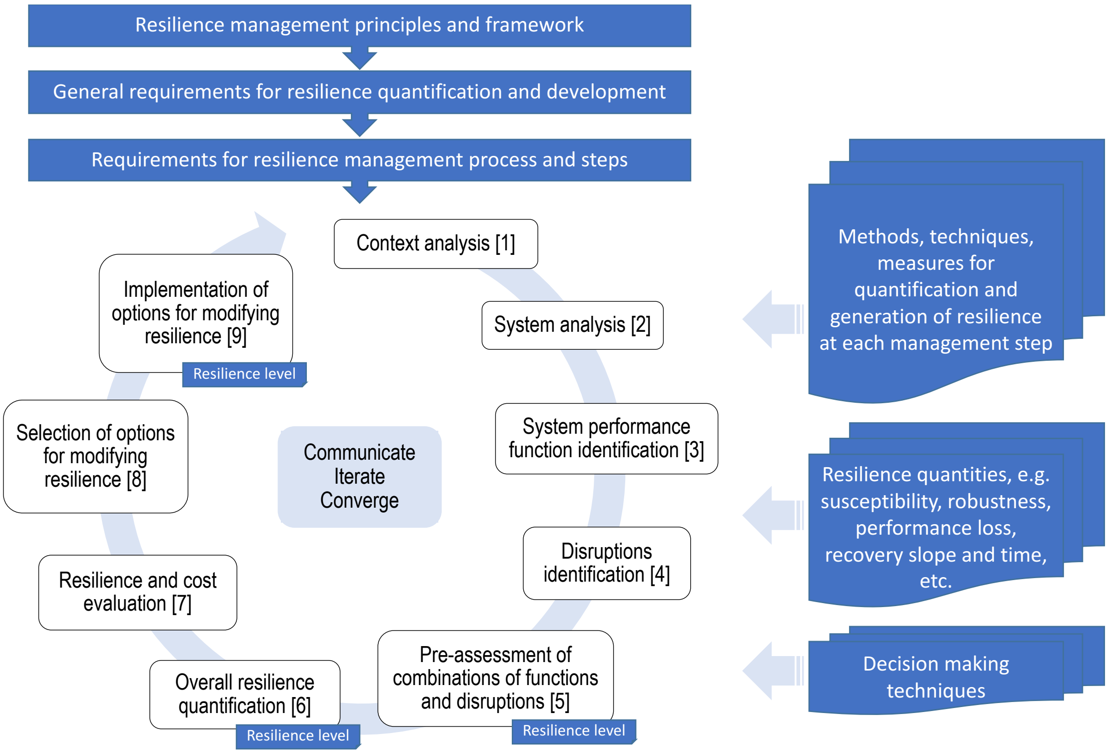

Figure 2.2: Generic resilience management process that consists of 9 steps and covers resilience quantification and development. The iterative process is governed by approved principles and framework, general requirements, specific process and steps requirements. Methods are used to support the approach in all steps (right side). Selected resilience quantities are used mainly in steps 5-9. Reproduced with permission

This resilience management process is iterative and consists of nine steps, which cover resilience quantification and development. The main idea is that the final choice of options for modifying resilience is made on the basis of the quantitative comparison of possible disruption events (losses) and of properties of a system itself (required investments in resilience). Governing principles, general and specific requirements should be developed. It would require a clear delineation of goals and scopes of the intended system. A taxonomy of methods is proposed in (Häring et al., 2017) **(3)**, however their application and selection of resilience quantities depend on the specified requirements and target level of resilience.

The idea that the resilience strategy depends on probable disruptions (stress- or risk-factors) appears to be well-recognized. It also illustrates interconnections and overlap between risk and resilience, which may lead to confusion and methodological inconsistencies in the two areas of expertise. To continue developing of the generic resilience management process, we juxtapose it to a standard risk management process, according to (ISO 31000:2009, E), figure 2.3.

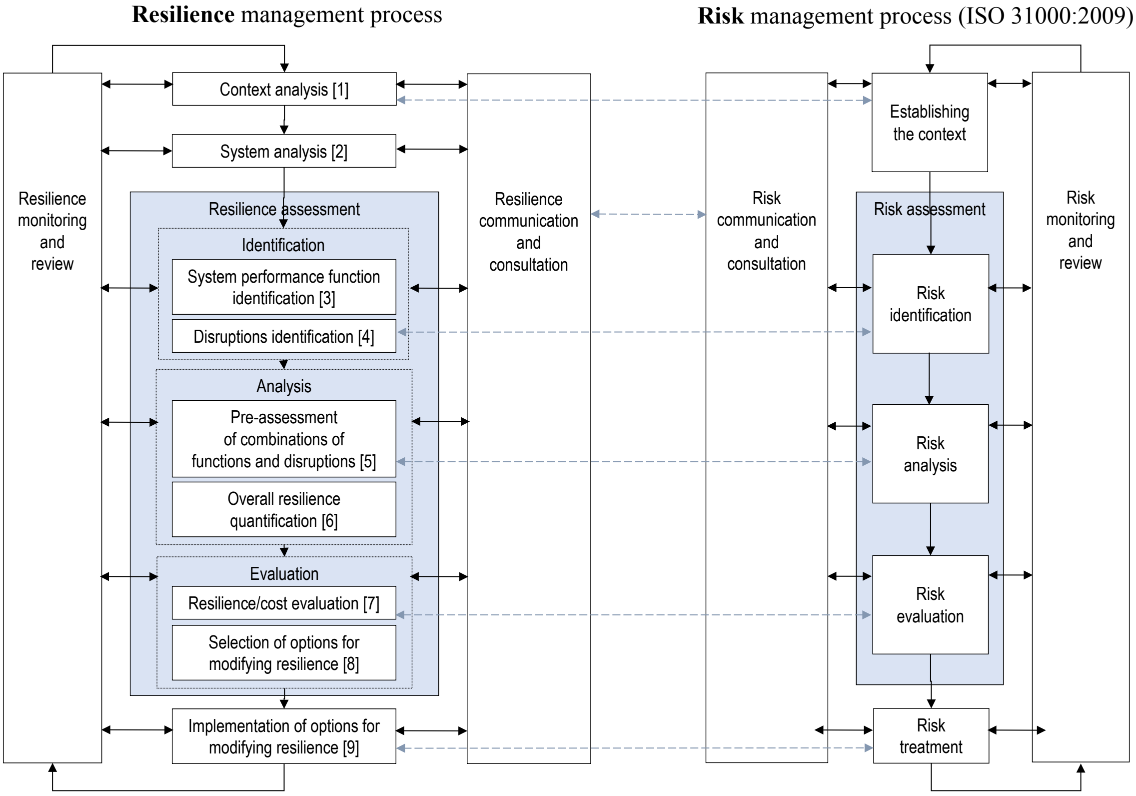

Figure 2.3: Juxtaposition of a resilience management process (left side) and a risk management process (right side). The risk management process is presented in accordance with the standard: *Risk management – Principles and guidelines* (ISO 31000:2009, E). Information flows between resilience and risk management processes at different steps are indicated by dashed arrows. Reproduced with permission

The nine-step resilience management process is well adapted to a standard risk-type management process, which is now a common practice across industries and countries. Accumulated experience in the risk field can be leveraged for a progressive build-up of resilience management. In fact, the development of a resilience management process should not become a goal in itself. It should be driven by business demands and integrated into an existing organizational structure. So, a tailored resilience management process can be applied selectively to the areas of high importance, for example to critical functions or sub-systems. It can be designed as an extension of risk management, or as an independent process. In the case of a separate risk and resilience management, transparency and barrier-free information exchange between these processes must be ensured.

A similar standard-inspired approach is proposed in (Heinimann and Hatfield, 2017). The originality of their framework is in framing resilience assessment and management concepts with 10 questions (deca-tuple set). This formulation is an alternative to the described nine-step process. Interestingly, the deca-tuple set relates to three classes of function: (i) biophysical, (ii) enabling, and (iii) cognitive. The latter includes state of awareness, anticipation, memory of past experience and adaptive individual behavior. In this context, the cognitive resilience function is indispensable for a resilient system, and makes a perfect transition to the next research topic developed in this thesis - decision theory.

References

(3) I. Häring, G. Sansavini, E. Bellini, N. Martyn, **Kovalenko, T.**, M. Kitsak, G. Vogelbacher, K. Ross, K. Bergerhausen, U. and Barker, and I. Linkov. Towards a generic resilience management, quantification and development process: General definitions, requirements, methods, techniques and measures, and case studies. In I. Linkov and J.M. Palma-Oliveira, editors, *Resilience and Risk: Methods and Application in Environment, Cyber and Social Domains*. Springer, Dordrecht, 2017

H.R. Heinimann and K. Hatfield. Infrastructure resilience assessment, management and governance – state and perspectives. In I. Linkov and J.M. Palma-Oliveira, editors, *Resilience and Risk: Methods and Application in Environment, Cyber and Social Domains*. Springer, Dordrecht, 2017

ISO 31000:2009(E). Risk management – Principles and guidelines. Standard, International Organization for Standardization, Geneva, 2009

3 Quantum decision theory parametrization

This chapter presents publication manuscripts that are either submitted, or are ready for submission. They are reprinted with permission from the co-authors. The copyright of the publication manuscripts is held by the respective copyright holders, see the following copyright notices.

Theory and Decision manuscript No.
(will be inserted by the editor)

Calibration of Quantum Decision Theory, aversion to large losses and predictability of probabilistic choices

Sabine Vincent * · Tatyana Kovalenko * ·
Vyacheslav I. Yukalov · Didier Sornette

Received: date / Accepted: date

Abstract We present the first calibration of Yukalov and Sornette's quantum decision theory (QDT) to a dataset of binary risky choice. First, we quantitatively account for the fraction of choice reversals between two repetitions of the experiment, using a probabilistic choice formulation in the simplest form without model assumption or adjustable parameters. The prediction of choice reversal is then refined by introducing heterogeneity between decision makers through their differentiation into two similar sized groups: "over-confident" and "contrarian". This supports the first fundamental tenet of QDT, which

---

S. Vincent
ETH Zurich, Department of Management, Technology and Economics, Scheuchzerstrasse 7, 8092 Zurich, Switzerland

T. Kovalenko
ETH Zurich, Department of Management, Technology and Economics, Scheuchzerstrasse 7, 8092 Zurich, Switzerland
Tel.: + 41 44 632 99 62
E-mail: tkovalenko@ethz.ch

V.I. Yukalov
ETH Zurich, Department of Management, Technology and Economics, Scheuchzerstrasse 7, 8092 Zurich, Switzerland
Bogolubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna 141980, Russia

D. Sornette
ETH Zurich, Department of Management, Technology and Economics, Scheuchzerstrasse 7, 8092 Zurich, Switzerland
Swiss Finance Institute, c/o University of Geneva, 40 blvd. Du Pont d'Arve, CH 1211 Geneva 4, Switzerland

* These authors contributed equally to this work.

models choice as an inherent probabilistic process, such that the probability of a prospect can be expressed as the sum of its utility and attraction factors. We propose to parameterise (a) the utility factor with a stochastic version of cumulative prospect theory (logit-CPT), and (b) the attraction factor with a constant absolute risk aversion (CARA) function. For this data set, and penalising the larger number of QDT parameters via the Wilks test of nested hypotheses, the QDT model is found to perform significantly better than logit-CPT at both the aggregate and individual levels, and for all considered fit criteria for the first experiment iteration and for predictions (second "out-of-sample" iteration). The distinctive QDT effect captured by the attraction factor is mostly appreciable for prospects with big losses. Our quantitative analysis of the experiment results supports the existence of an intrinsic limit of predictability, which is associated with the inherent probabilistic nature of choice.

Keywords Quantum decision theory · Prospect probability · Utility factor · Attraction factor · Stochastic cumulative prospect theory · Predictability limit

1 Introduction

The principal goal of decision theory is to understand and predict the choices of decision makers, in particular when the decisions involve risky options. "Classical" economists use the *Homo economicus* assumption that decision making is the deterministic process of maximising an expected utility (Bernoulli, 1738, von Neumann and Morgenstern, 1947, Savage, 1954). This formulation has been shown to lead to many paradoxes when confronted with real human decision makers. Accumulated empirical data reveal systematic behavioural patterns that indicate violation of the classical axioms. These violations include (a) common consequence and common ratio effects, which are inconsistent with the axiom of independence from irrelevant alternatives (Allais, 1953), (b) preference reversal phenomenon (Lichtenstein and Slovic, 1971, Lindman, 1971) that is associated with a failure of procedure invariance and the axiom of transitivity (Loomes and Sugden, 1983), and (c) framing effects as a breakdown of descriptive invariance (Tversky and Kahneman, 1974). Many models have been introduced to explain and predict observed cognitive and emotional biases (Camerer et al., 2003, Machina, 2008). A number of theories have been advanced, such as prospect theory (Edwards, 1955, 1962, Kahneman and Tversky, 1979), rank-dependent utility theory (Quiggin, 1982, 1993), cumulative prospect theory (Kahneman and Tversky, 1992), configural weight models (Birnbaum, 1974, 2008), regret theory (Loomes and Sugden, 1982, 1987), maximin expected utility model (Gilboa and Schmeidler, 1989), Choquet expected utility model (Gilboa, 1987, Schmeidler, 1989) and many others. However, various attempts to extend utility theory by constructing non-expected utility functionals do not avoid common pitfalls in modeling risk aversion (Safra and Segal, 2008), cannot in general resolve the known classical paradoxes such as the conjunction fallacy, disjunction effect, and were criticized for employing ambiguity aversion to rationalize Ellsberg choices (Al-Najjar and Weinstein, 2009). Moreover, extending the classical utility theory has been claimed "end[ing] up creating more paradoxes and inconsistencies than it resolves" (Ibid.).

One of the difficulties in modeling decision makers' behaviour is associated with the variability of their choices. There is compelling evidence from a substantial body of psychological and economic research that people are not only different in their preferences (corresponding to between-subject variability), but, importantly, they do not perform deterministic choices (and thus exhibit within-subject variability) (Mosteller and Nogee, 1951, Tversky, 1969, Hey, 2001). A person in a nearly identical choice situation on repeated occasions often opts for different choice alternatives, and the magnitude of choice probability variations is context dependent. Choice reversal (switching) rate has been reported between 20 and 30%, and for some tasks can be close to 50% (Camerer, 1989, Starmer and Sugden, 1989, Hey and Orme, 1994, Ballinger and Wilcox, 1997, Rieskamp et al., 2006, Regenwetter et al., 2011). Thus, at the aggregate and individual levels, decision makers do not seem to settle on the choice that exhibits the largest unequivocally defined desirability. To account for variability of individual choice, and to help formalise economic models, the previously mentioned (expected utility and non-expected utility) deterministic theories have been combined with stochastic components.

At an early stage, the development of probabilistic models of choice and preference was associated with psychophysics. Thurstone's law of comparative judgement (Thurstone, 1927) and Luce's choice axioms (Luce, 1959) imply models that are specimens of the two broad classes of probabilistic choice models.[1] Respectively, the classes are (Luce and Suppes, 1965, Marley, 1992, Rieskamp et al., 2006): (i) random utility models, which combine stochastic utility function with deterministic choice rule, i.e. the maximisation of a random utility at each repetition of a decision; and (ii) constant (fixed) utility models, which assume a fixed numerical utility function over the choice outcomes complemented by a probabilistic choice rule, i.e. response probabilities that are dependent on the scale values of the corresponding outcomes. For instance, cumulative prospect theory has been supplemented with the probit (Hey and Orme, 1994) or the logit choice functions (Carbone and Hey, 1995, Birnbaum and Chavez, 1997). Another class of models suggest the existence of (iii) a random strategy selection (or random preferences) such that, within each strategy (or preference state), both elements, utility and choice process, are deterministic. Random preference models (aka mixture models) assume probabilistic distribution of decision maker's underlying (latent) preferences, and interpret choices as if they are observations drawn from such a distribution

---

[1] For historical connections between Thurstonian model and Luce's choice model, see for example (Pleskac, 2012).

(Heyer and Niederee, 1989, 1992, Niederee and Heyer, 1997, Regenwetter, 1996, Regenwetter and Marley, 2001, Regenwetter et al., 2010, 2011, Loomes and Pogrebna, 2014, 2017). Different stochastic specifications has been explored, and a large literature has evolved (Marschak, 1960, Block and Marschak, 1960, Yellott, 1977, Iverson and Falmagne, 1985, Heyer and Mausfeld, 1987, Marley, 1968, 1989a,b, Luce and Narens, 1994, Harless and Camerer, 1994, Hey, 1995, Hey and Carbone, 1995, Luce, 1994, Ballinger and Wilcox, 1997, Loomes and Sugden, 1995, 1998, McFadden and Train, 2000, Fishburn, 2001, Loomes et al., 2002, Hey, 2005, Myung et al., 2005, Birnbaum, 2006, Rieskamp, 2008, Wilcox, 2008, Davis-Stober, 2009, Blavatskyy and Pogrebna, 2010, Conte et al., 2011, Regenwetter and Davis-Stober, 2012, Regenwetter et al., 2014, Mäs and Nax, 2016).

Summarising the above, the necessity of a stochastic approach for the modeling of choices is widely recognized. The need to prioritise the advancement of research concerned with probabilistic descriptions, as compared to the development of new versions of deterministic behavioural models, has been pointed out for example in (Hey and Orme, 1994, Hey, 2005, Rieskamp, 2008). In fact, the axiomatic expected utility theory, when extended to incorporate truncated random errors, has been demonstrated to explain experimental data at least as well as cumulative prospect theory (Blavatskyy, 2005). At the same time, we suggest that the *nature of the stochasticity of choices* deserves more attention, and some of the current interpretations may require reconsideration.

Firstly, one of the prevalent views in the literature is that the observed probabilistic choices are a result of the bounded rationality of decision makers. Empirically documented effects, such as preference reversal, similarity, compromise and attention effects, have often been classified as “inconsistencies” of people’s behaviour (Rieskamp et al., 2006), which is mistaken and noisy (Hey, 2005). In this interpretation, the core of the choice process is still deterministic, in the sense that the decision maker strives to choose the best alternative but, doing so, she makes errors either in the evaluation of the options (e.g. a measurement error (Hey and Orme, 1994)) or in the implementation of her choice (e.g. an application error with a constant probability of its occurrence (Harless and Camerer, 1994, Rieskamp and Otto, 2006)). The standard way of using such a stochastic approach is to assume a probability distribution over the values characterizing the errors made by the subjects in the process of decision making. Such stochastic decision theories can be termed as “deterministic theories embedded into an environment with stochastic noise”, and are typical of (i) random utility models and (ii) fixed utility models.

Another perspective is to consider that the stochastic elements are technical devices added to the deterministic theory to allow for its calibration to experiments, with the implicit or explicit understanding that the stochastic component of the choice may result from the component of the utility of a decision maker that is unknown or hidden to an observer trying to rationalize the choices made by the decision maker (Luce and Suppes, 1965, McFadden, 1974). This interpretation is relevant to models with (iii) random preferences. In this view, a probabilistic model accounts for the empirically observed behavioural inconsistencies, however their origin and causes are often put out of the scope of the discussion.

Finally, stochastic assumptions often remain implicit, though they play a defining role in the formulation of testable hypotheses and the selection of methods of statistical inference (Hey, 2005). Different probabilistic specifications have been shown to lead to possibly opposite predictions for the same core (deterministic) theory (Hey and Orme, 1994, Hey, 1995, Loomes and Sugden, 1995, Carbone and Hey, 2000, Loomes, 2005). These emphasize that "stochastic specification should not be considered as an 'optional add-on,' but rather as integral part of every theory which seeks to make predictions about decision making under risk and uncertainty" (p. 648) (Loomes and Sugden, 1995).

In our view, strong probabilistic theories, which assign a precise probability for each option to be chosen, provide valuable modeling tools. They should not be perceived as mere extensions of deterministic core theories. Rather, a general probabilistic framework that highlights the intrinsic stochastic origin of decision making should be put to the forefront. Arguably, among the classes named above, random preference models (mixture models) correspond the most to this approach (Loomes, 2015). Alternatively, models based on stochastic processes have been introduced to represent mental deliberation and account for choice and reaction time jointly, as well as to model (longitudinal) panel data. These include decision field theory (Busemeyer and Townsend, 1993), ballistic accumulator models (Brown and Heathcote, 2008), media theory (Falmagne, 1996, Falmagne and Ovchinnikov, 2002), sequential sampling models (Forstmann et al., 2016), stochastic token models of persuasion (Falmagne, 1997) and so on.

The quantum decision approach that we will present and test here resonates with this strand of research emphasizing that decision making might be intrinsically probabilistic. While there is a huge literature briefly mentioned above on probabilistic decisions, the prominent advantage of quantum decision theory is that it is by essence structurally probabilistic. In other words, the whole theoretical construction of how people make decisions cannot be separated from a probabilistic frame. Contrary to classical stochastic decision theory in economics, we do not assume that choices are deterministic, with just some weak disturbance associated with errors. In quantum decision theory, a probabilistic decision is not a stochastic decoration of a deterministic process: a random part is unavoidably associated with any choice, which can be interpreted as representing subconscious hidden neuronal processes. The difference between the classical stochastic decision theory in economics and quantum decision theory is similar to the difference between classical statistical physics and quantum mechanical theory. In the former, all processes are assumed to be deterministic, with statistics coming into play because of errors and statistical fluctuations, such as no precise knowledge of initial conditions and the impossibility of measuring exactly the locations and velocities of all particles. In contrast, quantum mechanics postulates that the precise states of particles are unknowable and, in the standard so-called Copenhagen interpretation, inherently so due to the essence of the laws of Nature. Similarly, the quantum decision theory used here embraces the view and actually requires in its very construction that decision making is intrinsically probabilistic.

There is a growing perception that the existence of probabilistic choices can be actually optimal in a certain broader sense. For instance, the occasional selection of alternatives that are dominated according to a particular desirability criterion, can actually be beneficial for an individual and/or a group when measured over large time scales. In evolutionary biology, a long-term measure of utility is known as reproductive value, which represents the expected future reproductive success of an individual. Natural selection favors those individuals, who behave as if maximising their reproductive value (Houston and McNamara, 1999). Similarly, traits such as "strong cooperation" (Henrich, 2004) and "altruistic punishment" (Fehr and Gächter, 2000a,b, Fehr and Fischbacher, 2003) are costly to the individual and do not seem to make sense from the perspective of a person's utility maximisation, but are selected in evolutionary agent-based models of competing groups in stochastic environments (Hetzer and Sornette, 2013a,b).

Stochastic decision making can provide an evolutionary advantage by being instrumental in overcoming adverse external and internal factors by:

- exploring uncertain complex environments with unknown feedbacks;
- discovering available choice options and variations of their utilities over time (McNamara et al., 2014);
- refining preferences by sampling and through comparative judgment (Stewart et al., 2006);
- learning using "trials and errors" and bridging a "description-experience gap" (Hertwig and Erev, 2009);
- adapting strategies at an individual and group levels, and introducing diversification.

Thus, choice variability should not be considered as an anomaly or exception. On the contrary, it may be an advantageous trait developed in humans, whose evolution is linked to a stochastic and uncertain environment. This view, incorporating the evidences reported in this paper, has been recently briefly summarised in (Sornette, 2017).

The quantum decision theory that we follow here was first introduced in (Yukalov and Sornette, 2008), with the goal of establishing an holistic theoretical framework of decision making. Based on the mathematics of Hilbert spaces, it provides a convenient formalism to deal with (real world) uncertainty and employs non-additive probabilities for the resolution of complex choice situations with interference effects. The use of Hilbert spaces constitutes the simplest generalization of the probability theory axiomatized by Kolmogorov (1956) for real-valued probabilities to probabilities derived from algebraic complex number theory. By its mathematical structure, quantum decision theory aims at encompassing the superposition processes occurring down to the neuronal level. This becomes especially important for composite (uncertain) measurements, with a formulation that differs from the diverse forms of probabilistic choice theory, including random preference models (mixture models), as the summary presentation of quantum decision theory in the appendix should help comprehend. Numerous behavioural patterns, including those causing paradoxes within other theoretical approaches, are coherently explained by quantum decision theory (Yukalov and Sornette, 2008, 2009, 2010, 2011, 2014, 2015a,b,c).

There are several alternative versions of quantum decision theory, which have been proposed in the literature, as seen for instance with the books (Khrennikov, 2010, Busemeyer and Bruza, 2012, Haven and Khrennikov, 2013, Bagarello, 2013) and the review articles (Yukalov and Sornette, 2009, Sornette, 2014, Busemeyer et al., 2014, Ashtiani and Azgomi, 2015), where citations to the previous literature can be found. The version of Quantum Decision Theory (henceforth referred to as QDT) developed by Yukalov and Sornette and used here principally differs from all other “quantum” approaches in two important aspects. First, QDT is based on a self-consistent mathematical foundation that is common to both quantum measurement theory and quantum decision theory. Starting from the von Neumann (1955) theory of quantum measurements, Yukalov and Sornette have generalized it to the case of uncertain or inconclusive events, making it possible to characterize uncertain measurements and uncertain prospects. Second, the main formulas of QDT are derived from general principles, giving the possibility of general quantitative predictions. In a series of papers, Yukalov and Sornette have compared a number of predictions with empirical data, without fitting parameters (Yukalov and Sornette, 2011, 2014, 2015b,c). This is in contrast with the usual way of constructing particular models for describing some concrete experiments, with fitting the model parameters from experimental data.

Until now, predictions of QDT were made at the aggregate level, non parametrically and assuming no prior information. This study intends to overcome these limitations, by developing a first parametric analytical formulation of QDT factors, enlarging the area of practical application of the theory and enabling higher granularity of predictions at both aggregate and individual levels.

For the first time, we engage QDT in a competition with decision making models, based on a mid size raw experimental data set of individual choices. The experiment was iterated twice (henceforth referred to as time 1 and time 2) and consists of simple choice tasks between two gambles with known outcomes and corresponding probabilities (i.e. binary lotteries). The data analysis reveals an inherent choice stochasticity, adding to the existing evidences, and supporting the probabilistic approach of QDT.

As a classical benchmark, we consider a stochastic version of cumulative prospect theory (henceforth referred to as logit-CPT) that combines cumulative prospect theory (CPT) with the logit choice function. Note that other models associated with "classical" theories, such as expected value (Pascal, 1670)[2] and expected utility theory (Bernoulli, 1738) are nested within it.[3]

Within QDT, a decision maker, who is exposed to several options, can choose any of these prospects with a certain probability. Thus, each choice option is associated with a prospect probability, which can be calculated as a sum of two factors: utility and attraction. In this paper, for the parametric formulation of QDT, we adopt the stochastic CPT approach (logit-CPT) for the utility factor, and incorporate a constant absolute risk aversion (CARA) into the attraction factor. This allows us to separate aversion to extreme losses and transfer it into the attraction factor.

We estimate parameters of the logit-CPT model and the utility factor of our QDT model with the hierarchical Bayesian method, as implemented in (Nilsson et al., 2011, Scheibehenne and Pachur, 2015, Murphy and ten Brincke, 2017), using identical data set as (Murphy and ten Brincke, 2017), which ensures straightforward model selection. The proposed QDT formulation is found to perform better at both aggregate and individual levels, and for all considered criteria of fit (time 1) and prediction (time 2). As expected, the most noticeable effect is achieved for prospects involving big losses, whereas the overall improvement is small on average.

The difficulty of achieving significant improvements in the prediction of human decisions, despite persistent attempts of different approaches, raises the question of the limit of predictability. We propose to rationalize quantitatively the limits of predictability of human choices in terms of the inherent stochastic nature of choice, which implies that the fraction of correctly predicted decisions is also a random variable. We thus propose a theoretical distribution of the individual predicted fractions, and compare it successfully to the experimental results.

To summarise, a first principal contribution of this article is to propose a parametric form of QDT that can be operationalized to allow for its parametric comparison with other models of decision making. Furthermore, the proposed formulation allows us to compare QDT to other models using individual rather than representative agent data.

---

[2] Blaise Pascal. Pensées. Republished several times, for instance 1972 in French by Le Livre de Poche, and 1995 in English by Penguin Classics, 1670.

[3] For review on tests of nested and especially non-nested hypotheses, see (Gourieroux and Monfort, 1994).

This article has the following structure. Section 2 presents empirical evidence supporting probabilistic choice frameworks. A simple nonparametric probabilistic model is proposed that can predict the frequency of preference reversals on the basis of the observed fraction of individuals making a choice in the first iteration of the experiment. Section 3 compares calibration and prediction results of the QDT model with the ones obtained for the stochastic model of CPT, both at the aggregate and individual levels. Section 4 investigates the limits of the improvement of choice predictions in the presence of the proposed probabilistic nature of decision making. Section 5 concludes.

2 Empirical evidence supporting probabilistic choice formulations

2.1 Basic experimental setting

Choice between gambles was called “the fruit fly of decision theory” (Kahneman and Tversky, 2000) as one of the simplest settings of choice under risk and elicitation of risk preferences. We consider a choice between two gambles A and B (i.e. binary lotteries), each of which consists of two outcomes, in a range from -100 to 100 monetary units (MU), with known probabilities that sum to one, as shown in table 1. Participants had to choose one of the lotteries, and were not allowed to express either indifference or lack of preference, thus a two-alternative forced choice (2AFC) paradigm was implemented. The experimental set included 91 pairs of static lotteries (i.e. outcomes and probabilities were not contingent upon a preceding choice of a decision maker) of four types: 35 pairs of lotteries with gains only; 25 pairs with losses only; 25 pairs of mixed lotteries with both gains and losses; and 6 pairs of mixed-zero lotteries with one gain and one loss and zero (status quo) as the alternative outcome. The first three types of binary lotteries cover the spectrum of risky decisions, while the mixed-zero type allows for measuring loss aversion separately from risk aversion (Rabin, 2000, Wakker, 2005). The set of lotteries was compiled from lotteries previously used in (Holt and Laury, 2002, Gaechter et al., 2007, Rieskamp, 2008). The collected empirical data of 142 participants (from the subject pool at the Max Planck Institute for Human Development in Berlin) was obtained from (Schulte-Mecklenbeck et al., 2016). Additional details of the experimental design, including a complete list of binary lotteries, can be found in (Murphy and ten Brincke, 2017), which exploits the same data set in their calibration of stochastic cumulative prospect theory (logit-CPT).

The experiment was repeated twice at an approximately two weeks interval (henceforth referred to as time 1 and time 2) with the same 142 subjects and the same set of 91 binary lotteries. At time 1, the order of lottery items and their spatial representation within a pair was randomized, and displayed in the reverse order at time 2. Consequently, the order and presentation effects

Table 1: Choice between two finite valued lotteries. If a decision maker chooses lottery A, then the outcome will be V_1^A with probability p_1^A, and V_2^A with probability $p_2^A = 1 - p_1^A$, and similarly if she chooses lottery B with the superscript changed from A to B. The outcomes can be either positive (gains) or negative (losses).

	Outcomes & Probabilities	
Lottery A	$(V_1^A; p_1^A)$ or $(V_2^A; p_2^A)$	$p_2^A = 1 - p_1^A$
Lottery B	$(V_1^B; p_1^B)$ or $(V_2^B; p_2^B)$	$p_2^B = 1 - p_1^B$

were mitigated. The experiment was incentive compatible with a two-part remuneration: a fixed participation fee, and a varying payment based on a randomly selected lottery from the choice set, which was played out at the end of both experimental sessions.

The recording of the choices between the same alternatives by the same subjects at two different times allows one to perform in-sample modeling (at time 1) and out-of-sample predictions (of time 2).

2.2 Analysis of the consistency and differences between times 1 and 2

2.2.1 Stability of the aggregate choice frequencies and variability of the individual preferences

Figure 1 compares the proportion of decision makers among the 142 subjects who chose option B at both time 1 and time 2 for each of the 91 binary lotteries. We refer to this proportion as the experimental "frequency" of choice B in a given pair of lotteries. As the diagonal in figure 1 represents what would be a perfect reproducibility of the choices at the two times, at the aggregate level, the first overall observation is that the frequency of the choice in each pair of lotteries is rather stable from time 1 to time 2, since the data points tend to cluster along the diagonal. The linear relationship shows that decision makers, as a group, exhibit a stable preference across time. The fact that the 91 lotteries sample essentially the full frequency interval $[0, 1]$ confirms that they cover a large set of preferences, from obvious gambles where one of the prospects is almost always preferred to more ambivalent gambles. The frequencies of the choices shown in figure 1 is a manifestation of the type of choices. It is also quite apparent that there is a significant scatter around the diagonal that signals a stochasticity in the revealed preferences of the 142 subjects.

The individual deviation of choices between times 1 and 2 is further quantified in figure 2, which plots the number of lottery pairs for which a given proportion of subjects have changed their choice. One can observe that individual choices of decision makers may vary significantly over time. In more than half of the binary lotteries, more than 30% of the subjects changed their answer between time 1 and time 2. The average rate of choice reversal (switching) per subject is slightly higher than 29%, which is in line with the values previously reported in the literature.

2.2.2 Quantitative rationalisation via probabilistic choices

The combined observation of the overall stability of the choices at the aggregate level (figure 1) and their variability at the individual level (figure 2) adds to the large body of empirical literature discussed in the introduction that purports that decisions are probabilistic rather than deterministic. However, it is interesting to test it quantitatively, as follows. For this, we propose a non-standard approach, which abstracts from any assumption on the probability model, algebraic core, and on the stimuli that promote the decisions. The only ingredient is to use the choices observed at time 1 as a measure of the corresponding prospect probabilities, without any fit. In other words, the frequency of a given choice over the population of decision makers is taken as a probe for the underlying probability for that choice, used in the usual frequentist interpretation of probabilities (Kendall, 1949). In a first step, this is done by assuming that all decision makers are described by the same unique probability for each choice. We take into account the sampling variabilities at times 1 and 2 by constructing confidence intervals for each frequency-based choice probability, using standard Bernouilli statistics.

Considering a given pair of lotteries, let us denote by X_t the event *"choosing lottery* $X \in \{A, B\}$ *at time* $t \in \{1, 2\}$*"*. For instance, if the decision maker chooses lottery A at time 1 and the lottery B at time 2, this is represented by the combined event $A_1 \bigcap B_2$. The overall stability of the choices at the aggregate level (figure 1) suggests the parsimonious assignment of a fixed stable probability p_j for each of the two choices in a given lottery pair j:

$$\mathbb{P}\left(A_{1,j}\right) = \mathbb{P}\left(A_{2,j}\right) = p_j \tag{1}$$

and

$$\mathbb{P}\left(B_{1,j}\right) = \mathbb{P}\left(B_{2,j}\right) = 1 - p_j \ . \tag{2}$$

This hypothesis consists in neglecting any heterogeneity between decision makers, thus assuming that they all have the same preference. Notwithstanding its simplicity, we now show that it is remarkably powerful at accounting for most of the observed shifts between times 1 and 2.

Indeed, because each choice among two lotteries within a pair is assumed probabilistic, this implies that repeating the experiment is expected to give possible choice shifts from A to B and vice-versa, just from the hypothesised probabilistic nature of the choice. Thus, the probability that a decision maker shifts her choice in a pair of lotteries is given by:

$$\mathbb{P}\left(\text{shift}\right) = \mathbb{P}\left(A_1 \bigcap B_2\right) + \mathbb{P}\left(B_1 \bigcap A_2\right) \ . \tag{3}$$

This expression conveys the fact that the shift could occur from the choice A at time 1 followed by the choice B at time 2. This is represented by $A_1 \bigcap B_2$. Or the decision maker might have chosen B at time 1 followed by the choice A at time 2. This is represented by $B_1 \bigcap A_2$. Considering both scenarios together leads to expression (3).

In the experiment, we deal with the same decision maker, facing the same set of two lotteries. Therefore, the successive decisions $A_1 \bigcap B_2$ or $B_1 \bigcap A_2$ are dependent because it is a repeated measure by design. However, let us assume that, when they form their choice at time 2, decision makers have forgotten their choices performed at time 1 (which is likely in the experimental set-up as the two iterations – time 1 and time 2 – were conducted approximately 2 weeks apart and the choice orders have been randomised). In the framework where their decisions are solely and completely captured by equations (1,2) expressing an intrinsic probabilistic choice structure, for a pair of lotteries we have $\mathbb{P}\left(A_1 \bigcap B_2\right) = \mathbb{P}\left(B_1 \bigcap A_2\right) = p(1-p)$, yielding

$$\mathbb{P}\left(\text{shift}\right) = 2p\left(1-p\right) \ . \tag{4}$$

In order to test the validity of prediction (4) on the experimental data, as mentioned above, we assume that the frequency of the most common choice for a given lottery pairs over the ensemble of all decision makers is a proxy for the probability p_j. Indeed, the frequency of the most common choice for a given pair j of lotteries gives an estimate of the so-called frequentist definition of the corresponding probability (Kendall, 1949), which converges to the true probability, if it exists, in the limit of very large samples. Similarly, we identify the probability $\mathbb{P}\left(\text{shift}\right)$ of a choice shift between times 1 and 2 with the proportion of decision makers having changed their choice between times 1 and 2. This prediction (4), which has no adjustable parameters, is shown as the blue smoothed continuous curve in figure 3, which plots the proportion of decision makers having changed their choice between times 1 and 2 as a function of the frequency of the most common choice at time 1. We note that it is easy to account for the sampling variabilities at times 1 and 2 by constructing confidence intervals for each frequency-based choice probability, using standard Bernouilli statistics. The corresponding confidence interval is presented in figure 5, which uses a slightly more refined model explained below.

Figure 3 shows that the main dependence is rather well captured by prediction (4), which we stress again is not a “fit” as there is no adjustable parameter. Expression (4) has a simple intuitive interpretation: clear-cut choices associated with large p_j's are aligned with strong and well-defined preferences, so that it is quite unlikely that a decision maker will change her choice; in contrast, when the frequency at time 1 for choosing a given lottery is close to even between the two lotteries, the decision makers are very likely to shift their choice at time 2. While these tendencies are obvious, what is less evident is the fact that the simple logical step leading to expression (4) accounts surprisingly well for the data, with no adjustment.

2.2.3 Evidence of heterogeneity between decision makers: a parsimonious description

While the agreement between data and prediction shown in figure 3 is remarkable, given that the prediction has no adjustable parameters, it is also clear that the model over-estimates the number of decision shifts as the data tends to be systematically below the theoretical prediction, in particular for the pairs of lotteries with close ties, i.e. for which decision makers show a large heterogeneity of choices and the proportion choosing the most frequently chosen lottery is not much above 50%. More precisely, for more frequently chosen options (with frequency of the most common choice above 75%), the observed frequencies are closer to the theoretical prediction, while, for less frequently chosen options, the deviation is larger. This can explain the bimodal structure of the histogram in figure 2.

In order to arrive at prediction (4), we have used two main assumptions: (i) the choices between times 1 and 2 are made as if a single probability describes each of them and (ii) the decision makers' preferences are homogenous, so that the same single probability $\{p_i, i = 1, ..., 91\}$ for each of the 91 pairs of lotteries characterises the full set of 142 subjects. We propose to keep the first assumption as part of a minimalist approach. As discussed briefly above, the second assumption flies in the face of enormous empirical evidence supporting the proposition that human decision makers exhibit significantly different risk preferences. This is particularly relevant to our discussion since the choices between the pairs of lotteries are specifically sensitive to the different levels of risk (as well as payoffs) associated with the competing lotteries in each pair.

Relaxing the assumption that all decision makers are identical can immediately be seen to help removing the discrepancy observed in figure 3. Indeed, consider the simplest situation generalising homogeneity, which consists in assuming the presence of two groups $i \in \{1, 2\}$ of decision makers of size $142F$ and $142(1-F)$ respectively (with $0 < F < 1$), for which $P(A^1_{j,1}) = P(A^1_{j,2}) = p_1$ and $P(A^2_{j,1}) = P(A^2_{j,2}) = p_2$, where $A^i_{j,t}$ is the most frequent choice in a given lottery pair j at time t by group i. Then, the aggregate probability of shift is

$$2Fp_1\,(1-p_1) + 2(1-F)p_2\,(1-p_2) \tag{5}$$

which is always smaller than its homogenised version (4) with the aggregate choice probability

$$p = p_1 F + p_2\,(1-F)\ \ . \tag{6}$$

This results from the concavity of the function $f(p) = p(1-p)$. In the case $F = 1/2$, this is also straightforwardly seen from the inequality $\left(p_1^2 + p_2^2\right)/2 \geq p_1 p_2$. The equality between expression (5) and (4) with (6) is recovered obviously for the homogeneous case, i.e. for $F = 0$ or $F = 1$.

We now propose a simple quantitative model by assuming the following ansatz for p_1 and p_2:

$$\begin{cases} p_1 = p + \alpha p\left(1-p\right) & \alpha \in [0,1] \\ p_2 = p - \beta p\left(1-p\right) & \beta = \alpha F / \left(1-F\right) \in [0,1] \end{cases} \tag{7}$$

where the value for β derives from (6). Intuitively, the ansatz $p_1 = p + \alpha p\left(1-p\right)$ in (7) states that the first group of decision makers tends to overweight the majority choice when the two lotteries are difficult to tell apart (region of p not too much larger than 1/2). We can refer to this first group as "over-confident". The second ansatz $p_2 = p - \beta p\left(1-p\right)$ in (7) states that the second group of decision makers tends to dislike the average preferred choice, the more difficult it is to decide between two lotteries. We call this second group "contrarian".

Calibrating this model (7) to the data shown in figure 3 by iterated tabu searches (Glover, 1993), we obtain the best estimates $\alpha = \beta = 1$ and $F = 0.5$, leading to the best model expressed from (7) as

$$\begin{cases} p_1 = 2p - p^2 \\ p_2 = p^2 \ , \end{cases} \tag{8}$$

which is represented in figure 4. The decision makers referred to as "over-confident" tend to exhibit much less uncertainty towards the most common choice. In contrast, the decision makers that we call "contrarian" tend to weaken or even oppose the most common choice.

Figure 5 presents the same data as figure 3 but the model is now taking into account the heterogeneity among decision makers via the simple ansatz (8) of two groups, "over-confident" and "contrarian". While this model is clearly over-simplified, it provides an excellent fit to the data confirming that, within the probabilistic choice framework, heterogeneity among decision makers is sufficient to account quantitatively for the observed changes of behaviour between times 1 and 2. The grey band represents the 90% confidence interval, which is delineated by the 5% and 95% quantiles, i.e. the area where 90% of the shifts should fall according to Monte Carlo simulations using the above model with two groups (3000 simulations per pairs of lotteries). This allows us to quantify the uncertainty band resulting from sampling variabilities at times 1 and 2, using standard Bernouilli statistics.

3 Calibration of quantum decision theory

3.1 Brief presentation of stochastic cumulative prospect theory (logit-CPT) and quantum decision theory (QDT)

Based on experiments in which 142 decision makers made 91 choices at two different times with the same set of choices but presented in different orders, the previous section has shown that the hypothesis that decisions are probabilistic provides a parsimonious and quantitative description of decision making. We thus endeavour to test two probabilistic choice theories, (i) stochastic cumulative prospect theory (logit-CPT) and (ii) quantum decision theory. Both theories are summarised in the Appendix.

Prospect theory (Kahneman and Tversky, 1979, 1992) is now the most famous alternative to expected utility theory. The outcomes are quantified through a value function v, weighted by subjective probabilities obtained from the objective probability via a non-additive weighting function w. Moreover, the value function separates gains and losses, where the notions of gains and losses are defined with respect to a reference point, here assumed to be zero. Cumulative prospect theory (CPT) can be combined with a probabilistic choice function, allowing for probabilistic deviations from the option that maximises the choice criterion with respect to alternative options. There are many probabilistic extensions of CPT, some of which are modeling something entirely separate from response errors using polyhedral combinatorics, such as, e.g. in (Regenwetter et al., 2014). The probabilistic version of CPT that we use here is called logit-CPT because the probability weighting scheme uses the logit function (see Appendix and below). Such stochastic extension is often perceived as an add-on to an intrinsically deterministic CPT approach that is necessary to account for the observed stochasticity of human choices, interpreted as errors or unobserved components of an underlying deterministic process.

Quantum decision theory (QDT) is based on two essential ideas: (a) an intrinsic probabilistic nature of decision making and (b) a generalisation of probabilities using the mathematics of Hilbert spaces that naturally account for entanglement between choices (Yukalov and Sornette, 2008, 2009, 2010, 2015a). Thus, in contrast to logit-CPT, it places the probabilistic nature of choice at the center of its construction. As recalled in the Appendix (see expressions (30-32)), a fundamental result of QDT is that the probability $p\left(\pi_n\right)$ of a given prospect π_n can in general be decomposed as the sum of two terms according to

$$p\left(\pi_n\right) = f\left(\pi_n\right) + q\left(\pi_n\right) \ . \tag{9}$$

The first term $f(\pi_n)$ is associated with the utility of the prospect under consideration and, therefore, is called the *utility factor*. The second term $q(\pi_n)$ accounts for interference and entanglement between prospect and state of mind, and results technically from the complex quantum nature of the probabilities describing the choices of decision makers. In decision theory, it characterizes subjective and subconscious processes of the decision maker related to other available prospects, as well as past experiences, beliefs and momentary influences, and is referred to as the *attraction factor*. We interpret the attraction factor as representing a subconscious attraction of a person to a given prospect. The attraction depends on the state of mind that can be influenced by external (i.e. situational) and/or internal (i.e. hunger, mood, fatigue, etc.) factors. For more precise definitions of the attraction factor, we refer to the appendix and to (Yukalov and Sornette, 2008, 2009, 2010, 2015a).

By the quantum-classical correspondence principle, when the quantum term $q(\pi_n)$ becomes zero, the quantum probability reduces to the classical probability, so that $p(\pi_n) \to f(\pi_n)$ for $q(\pi_n) \to 0$, with the normalization $\sum_n f(\pi_n) = 1$ with $0 \leq f(\pi_n) \leq 1$. In the sequel, we use a logit-CPT form for the utility factor $f(\pi_n)$ given by expression (16) below, which corresponds to the first term in equation (15). We assume that logit-CPT can adequately characterize the utility of an isolated prospect for a decision maker. While logit-CPT incorporates some subjective deviations of values and probabilities, it treats each prospect separately, with no interference between the different prospects or no interference between a given prospect and the state of mind.

As already mentioned, the attraction factor embodies the additional complex unconscious deliberations and preferences associated with decision making. By construction, it enjoys the following properties (Yukalov and Sornette, 2008, 2009, 2010, 2015a). It lies in the range $-1 \leq q(\pi_n) \leq 1$ and satisfies the *alternation law* $\sum_n q(\pi_n) = 0$. In addition, for a large class of distributions, there exists the *quarter law*

$$\frac{1}{N} \sum_{n=1}^{N} |q(\pi_n)| = \frac{1}{4} \; . \tag{10}$$

In the presence of two competing prospects, one can show that, in the absence of any other information (the so-called "non-informative prior"), one obtains

$$|q(\pi_n)| \approx 0.25 \; , \tag{11}$$

which makes it possible to give quantitative predictions in absence of additional information (Yukalov and Sornette, 2011, 2014, 2015b,c). In the following, we go beyond (11) and introduce a mathematical expression (49) with constant absolute risk aversion (CARA) utility function (50) for the attraction factor, which corresponds to the second term in equation (13) and is motivated by the structure of the pairs of lotteries presented to the decision makers.

3.2 Methodology to estimate logit-CPT and QDT

We follow and extend the procedure of parameters estimation proposed by Murphy and ten Brincke (2017). We first summarise their method and then extend it to QDT.

According to stochastic decision theories such as logit-CPT, the option A_j of the pair j of lotteries is chosen by a subject over the option B_j with a probability p_{A_j}, which depends on individual parameters. These parameters can be estimated by fitting the model to the data obtained at time 1 and then used for predicting the outcomes at time 2.

The answers from the decision maker $i \in \{1 \dots 142\}$ at time 1 are denoted $\left(\Phi^i_j\right)_{j=1}^{91}$.

$$\Phi^i_j = \begin{cases} 0 \text{ if subject } i \text{ chooses } A \text{ in the } j^{\text{th}} \text{ gamble} \\ 1 \text{ if subject } i \text{ chooses } B \text{ in the } j^{\text{th}} \text{ gamble} \end{cases} \tag{12}$$

Given the choices $\left(\Phi^i_j\right)_{j=1}^{91}$, the individual parameters of the decision maker i can be estimated with a maximum likelihood method. A natural choice for the objective function is

$$\Pi^i = \prod_{j=1}^{91} p_{A_j}^{1-\Phi^i_j} p_{B_j}^{\Phi^i_j} \tag{13}$$

However, it has been shown by Nilsson et al. (2011) that this optimization method gives unreliable estimates at the individual level, since a shift of a single answer sometimes leads to very different parameters estimates. The *hierarchical maximum likelihood* method based on the work of Farrell and Ludwig (2008) fixes this issue by introducing the assumption that the individual parameters are distributed in the population with a given density distribution. The optimization is then performed for each subject, weighting the objective functions with the density distributions obtained at the population level. Murphy and ten Brincke (2017) applied this method to the experimental data described in section 2.1. Applied to stochastic CPT briefly described in the Appendix, the distributions of the parameters α, λ, γ and δ were assumed to be lognormal. Each log-normal distribution is defined through its location parameter μ and its scale parameter σ, which were estimated with a maximum likelihood method at the aggregate level.

The exact same data and parameters estimation procedure were used in the analysis of the present article, which allows for a direct comparison of stochastic cumulative prospect theory and quantum decision theory. For stochastic cumulative prospect theory, we are able to recover precisely the quantitative results reported by Murphy and ten Brincke (2017). In other words, we did not use the parameters reported by Murphy and ten Brincke (2017) but re-estimated them ourselves completely independently, reproducing entirely the whole calibration procedure for the logit-CPT. Then, we extended the procedure to calibrate and test QDT as explained below. The detailed description of the methodology follows.

- *At the aggregate level*
At the aggregate level, the parameters are estimated with a maximum likelihood method for both models (logit-CPT and QDT). The objective function is

$$\Pi^{\text{agg}} = \prod_{i=1}^{142} \prod_{j=1}^{91} p_{A_j}^{1-\Phi^i_j} p_{B_j}^{\Phi^i_j} \tag{14}$$

where the probability of choosing option A over option B is defined as follows (see Appendix):

QDT:

$$p_{A_j} = \frac{1}{1 + e^{\varphi\left(\tilde{U}_{B_j} - \tilde{U}_{A_j}\right)}} + \min\left(f_{A_j}, 1 - f_{A_j}\right) \tanh\left(a\left(U_{A_j} - U_{B_j}\right)\right) \quad (15)$$

logit-CPT:

$$p_{A_j} = \frac{1}{1 + e^{\varphi\left(\tilde{U}_{B_j} - \tilde{U}_{A_j}\right)}} \,. \quad (16)$$

To be clear, associated with the utility factor, $\tilde{U}$ represents the utility according to the CPT framework defined by expression (43), while U, which is defined by expression (50) as the CARA function with a coefficient of absolute risk aversion η, enters into the definition (49) of the attraction factor.

Note that the QDT formulation has two additional parameters (a and η) compared to logit-CPT, so that the later is nested in QDT (it is retrieved from the QDT formulation by setting $a = 0$).

- *At the individual level*

When applied finally to the individual level, the parameters are estimated with a hierarchical maximum likelihood method for both models (logit-CPT and QDT). In a nutshell, this means first estimating the distribution of parameters at the aggregate level to obtain prior distributions, which are then used as weights penalising possible over-determinations at the individual level. The objective function for each subject i is

$$\Pi^i = g_\alpha g_\lambda g_\gamma g_\delta \prod_{j=0}^{91} p_{A_j}^{1-\Phi_j^i} p_{B_j}^{\Phi_j^i} \quad (17)$$

where

- g_X is the distribution of the parameter $X \in \{\alpha, \lambda, \gamma, \delta\}$, according to the experimental results from (Murphy and ten Brincke, 2017),
- and the probabilities are for QDT:
  $p_{A_j} = \frac{1}{1+e^{\varphi\left(\tilde{U}_{B_j} - \tilde{U}_{A_j}\right)}} + \min\left(f_{A_j}, 1 - f_{A_j}\right) \tanh\left(a^{\text{agg}}\left(U_{A_j}^{\text{agg}} - U_{B_j}^{\text{agg}}\right)\right)$.
  Note that the exponent “agg” indicates that, at the individual level, a and η are not seen as parameters, but replaced by their optimal values found at the aggregate level.
- For logit-CPT, $p_{A_j} = \frac{1}{1+e^{\varphi\left(\tilde{U}_{B_j} - \tilde{U}_{A_j}\right)}}$

In particular, at the individual level, the QDT formulation involves the same number of individual parameters as the logit-CPT formulation.

The solver used for all the optimizations is the *fminsearch* function from MATLAB (Nelder&Mead simplex algorithm), the starting values of the parameters are chosen with a tabu search.

3.3 Calibration and prediction at the aggregate level

First, a cautionary remark is in order. If individuals satisfy logit-CPT with different parameter values, then the aggregate will violate logit-CPT in virtually any scenario. It would thus seem that the mixing and matching of individual and aggregate models and data are misleading. But, as explained above, the hierarchical Bayesian logit-CPT implementation (Nilsson et al., 2011, Scheibehenne and Pachur, 2015, Murphy and ten Brincke, 2017) is only used to provide not unreasonable priors for the parameters of logit-CPT at the individual levels. There is no normative statement about the applicability of the logit-CPT at the aggregate level. This exercise should be just considered as a convenient procedure to obtain more robust estimations at the individual levels, as shown by previous works mentioned above.

At the aggregate level, the optimization problem for QDT involves seven parameters: five for the utility factor (formula (46)) and two for the attraction factor (formula (49)). The logit-CPT model is nested in the QDT one (null hypothesis: $a^{\mathrm{agg}} = 0$) (Gourieroux and Monfort, 1994), which implies that one has to be very careful with choosing a statistical test so that it can “punish” the more general formulation. An often used method is to invoke the Akaike Information Criterion to penalise the larger number of parameters of QDT versus logit-CPT, in order to compare them against each other properly. In fact, the likelihood ratio-test (also known as the Wilks test) (Wilks, 1938) is the most powerful test for nested hypothesis and superseded the Akaike Information Criterion for nested tests. For nested hypotheses, one can show that two times the log-likelihood ratio has a chi-square distribution with a number of degrees of freedom equal to the difference in the number of parameters between QDT and logit-CPT (which is 2, a and η), under the null that the generating process is the logit-CPT (i.e., the model with the smaller number of parameters). Performing the test, we find that the likelihood ratio-test rejected the null hypothesis (logit-CPT) at the 95% level. In other words, the logit-CPT is insufficient to describe the data and the QDT formulation is providing a significant improvement, which is sufficiently large to compensate for the “cost” of an additional parameter.

Table 2 shows the values of the parameters for the two parametrisations. The CARA utility function U with the obtained parameters of the attraction factor is illustrated in figure 16 in the appendix. In particular, since $|q|$ depends on the difference $U_A - U_B$, the attraction factor is small except for some gambles involving big losses. The attraction factor thus accounts for the observation that, in experiments, people do not care much about medium payments, but respond to large losses.

Moreover, most of the parameters describing the utility term (α, γ, δ and φ) of QDT are close to the ones obtained with logit-CPT. However, the kink of the CPT value function at 0 quantified by λ is smaller for QDT: this means that, though loss might loom more than gain in general ($\lambda > 1$), this effect is significantly transferred to a risk aversion for big losses that is incorporated in the attraction factor ($q \neq 0$) within QDT.

Table 2: Estimated values of the parameters for the two models (logit-CPT and QDT). The values found for the QDT model are close to those obtained for logit-CPT for most of the common parameters (α,δ,γ,φ). The loss aversion parameter λ is smaller with QDT (but still larger than 1) because aversion to large losses is captured by the additional parameters of the QDT attraction factor (a and η).

	α^{agg}	λ^{agg}	δ^{agg}	γ^{agg}	φ^{agg}	a^{agg}	η^{agg}
logit-CPT	0.73	1.11	0.88	0.65	0.30	-	-
QDT	0.69	1.02	0.89	0.63	0.37	1.5	0.05

Figure 6 demonstrates that the QDT model reduces a lot the prediction errors for pairs of lotteries involving big losses (mixed-lotteries and lotteries with only losses). Though, for other lotteries, the improvement might not seem significant, table 3 shows that QDT reduces the residual sum of squares when summed over all gambles, and also when summed separately for each type of gambles (only losses, only gains and mixed-gambles). In other words, due to the fact that QDT predicts large risk aversion to big losses and only moderate risk-aversion to small losses, QDT outperforms logit-CPT in predicting choices of gambles with big losses.

Table 3: Statistics of the calibrations (time 1) and predictions (time 2) for both parametrisations (logit-CPT and QDT) at the aggregate level. The residual sum of squares (RSS) is smaller with QDT for both times and separately for each type of gambles (only losses, only gains and mixed-gambles). The correlation is closer to 1 with QDT.

		logit-CPT	QDT
RSS for all gambles:	FIT	0.73	0.52
	PREDICTION	0.76	0.59
RSS for gambles with only losses:	FIT	0.22	0.15
	PREDICTION	0.26	0.13
RSS for mixed-gambles:	FIT	0.27	0.17
	PREDICTION	0.21	0.18
RSS for gambles with only gains:	FIT	0.24	0.21
	PREDICTION	0.29	0.28
Correlation:	FIT	0.93	0.95
	PREDICTION	0.93	0.95

3.4 Calibration and prediction at the individual level

At the individual level, since the two formulations include the same number of parameters, the model selection can be done according to the log-likelihoods (in this case, BIC is equivalent to using the Akaike Information Criterion -

AIC due to the same number of parameters): the preferred model is the one that has the largest log-likelihood. According to this criterion, we find that the QDT model has the highest predictive power (see table 4).

Table 4: Model selection (time 1) according to the log-likelihood criterion. Since the two models have the same number of parameters, this gives the same result as a selection based on the AIC. The log-likelihood is larger for QDT for most subjects and on average, so the QDT model is preferred to logit-CPT.

	logit-CPT	QDT
Proportion of subjects best predicted by	34.51%	65.49%
Mean of the log-likelihood:	-86.53	-85.77

In particular, the QDT model is selected when the average log-likelihoods are compared, and also for most subjects (65.49%) when the selection is performed individually. Figure 7 provides a comparison of the log-likelihoods obtained with logit-CPT and QDT for all the subjects.

Moreover, table 5 highlights that the averages of the explained fractions, predicted fractions, and log-likelihoods at both times are slightly larger for the QDT model compared with the logit-CPT formulation. While the improvements of these diagnostics obtained with QDT over logit-CPT are not large, they are of the same size as those obtained by Murphy and ten Brincke (2017) in their evaluation of different competing models (excluding QDT). In section 4, we propose an explanation for these results, based on an intrinsic limit of predictability associated with the intrinsic probabilistic nature of decision making.

Table 5: First row: average explained and predicted fractions of choices among decision makers for the logit-CPT and QDT models. Second row: average loglikelihoods. The results obtained with the QDT model are slightly better than with stochastic cumulative prospect theory for the explained and predicted fractions and the log-likelihoods at both times.

		logit-CPT	QDT
Explained fractions:	FIT	0.76	0.77
	PREDICTION	0.73	0.74
Loglikelihood:	FIT	-86.53	-85.77
	PREDICTION	-99.31	-98.33

A closer look at the predicted fractions of choices for each pair j of lotteries (figure 8) reveals that the improvement obtained with QDT for the average predicted fraction is especially noticeable in some gambles including big losses. For those particular gambles, the quantum attraction factor is very significant. For the other gambles, the predictions are of the same quality with both methods.

The individual parameters obtained for the utility factor tend to differ by less than 10% from those obtained with logit-CPT (see figure 9): this implies that, for lotteries with negligible attraction, QDT gives individual predictions that are close to those given by logit-CPT.

3.5 Hints for the need of a multi-modal extension of QDT

As recalled in the Appendix (see expressions (30-32)), and as given by expression (9), a fundamental result of QDT is that the probability of a given prospect π_n can in general be decomposed as the sum of two terms, the utility factor $f(\pi_n)$ (which is a stable individual trait of the decision maker) and the attraction factor $q(\pi_n)$ (which is dynamic, changing, and state-dependent). We have used the logic-CPT for the utility factor and the mathematical expression (49) with the CARA function (50) for the attraction factor.

This formulation and our tests have assumed that we can make generalisations for a population. Comparing with logit-CPT with the same assumption of homogeneity, our calibration tests have supported the usefulness of the QDT formulation in the form of an added value brought by the attraction factor. However, section 2.2.3 has shown that there is strong evidence for two main groups of decision makers, the over-confidents and the contrarians. This hypothesis provided a remarkable good fit shown in figure 5, obtained with model (8) represented in figure 4.

As can been seen from (8) and figure 4, it is interesting to notice that the choice probabilities p_1 and p_2 of the over-confidents and contrarians are such that $p_1(p \simeq 0.5) = p{+}0.25$ and $p_2(p \simeq 0.5) = p{-}0.25$ in a rather large domain of p values from $p = 0.5$ up to not to close to 1. The ±0.25 terms can be interpreted as attractor factors in a QDT formalism, which allows one to account for different risk aversions among decision makers. The relation $p_1(p \simeq 0.5) = p + 0.25$ for the over-confidents corresponds to the non-informative prior (11) for the attraction factor, with a positive sign expressing a group of decision makers who are over-optimistic about the value of their choices. The relation $p_2(p \simeq 0.5) = p - 0.25$ for the contrarians corresponds to the non-informative prior (11) for the attraction factor, with a negative sign expressing a group of decision makers who are distrusting the average choices.

These considerations suggest novel directions to develop further QDT, in which the sign and amplitude of the attraction factor is not just determined by the structure of the decision problem but also by the state of mind of the decision makers. This will be developed in future works.

4 Limits of predictability with probabilistic choices

We return to the considerations and tests of section 2 that strongly suggest that decisions are probabilistic rather than deterministic. We test further this hypothesis and show that it allows us to quantitatively account for the limits of predictability observed in the experiments.

Indeed, table 5 showed that the current analytical formulation of QDT allowed us to improve the individual fit and prediction for most subjects and on average, but with a rather small improvement of prediction on average, going from 73% for logit-CPT to 74% for QDT. The same issue was encountered by Murphy and ten Brincke (2017) who found that, while their implementation of the hierarchical maximum likelihood method improved the reliability of the parameter estimates and the log-likelihoods of results at time 2, the average predicted fraction did not improved compared with the one obtained with the usual maximum likelihood estimation method. This hints at a hard "barrier" preventing to improve further the fraction of decisions. Actually, if choices are probabilistic, this barrier obtains a natural explanation.

4.1 Distribution of the predicted fractions

For a given pair of lotteries $j \in \{1 \dots N\}$ and a given decision maker i, we define the probability $p^i_{A_j}$ with which the lottery A is picked over B. Likewise, a probability $p^i_{B_j}$ is defined, and $p^i_{B_j} = 1 - p^i_{A_j}$.

Suppose that the probabilities $p^i_{A_j}$ and $p^i_{B_j}$ are known and stable in time. Then the best prediction for the pair of lotteries j is to assume that the decision maker will prefer the most likely choice. Consequently, the choice regarding lotteries of the pair j can be seen as a Bernoulli trial, with a probability of success p^i_j larger than 0.5:

$$p^i_j = \max\left(p^i_{A_j}, p^i_{B_j}\right) \tag{18}$$

Let P^i be the fraction of choices predicted correctly for subject i. P^i correspond to the fraction of successes in a sequence of N independent Bernoulli trials with different probabilities of success. Thus the random variable P^i follows a Poisson binomial distribution.

Given the success probabilities $\left(p^i_j\right)_{j\in\{1\dots N\}}$, the discrete distribution can be numerically approximated using a discrete Fourier transform (Fernández and Williams, 2010) by the following formula:

$$\begin{cases} \mathbb{P}\left(P^i = k/N\right) = \dfrac{1}{N+1}\displaystyle\sum_{l=0}^{N} C^{-lk} \prod_{m=1}^{N} \left(1 + \left(C^l - 1\right) p^i_j\right) \qquad k \in \{0 \dots N\} \\ C = \exp\left(\dfrac{2\omega\pi}{N+1}\right) \end{cases} \tag{19}$$

where ω stands for the pure imaginary number such that $\omega^2 = -1$.

For the experiment described in section 2.1, the theoretical Poisson binomial distributions of the predicted fraction of choices for a group of typical decision makers are plotted in figure 10. For these distributions, individual prospect probabilities of the most likely choice ($p_j^i > 0.5$) for each of the 91 pairs of lotteries j are estimated with the QDT model at time 1. These values are then inserted in expression (19) to explain (“in-sample) at time 1 and predict (“out-of-sample”) at time 2 the fraction of correct choices (“correct” in the sense that the choice corresponds to the probability larger than 0.5 as estimated by the QDT calibration). The group of typical decision makers (7 subjects) is chosen such that the mode of their theoretical Poisson binomial distribution P^i is equal to 0.77, i.e. the median value among the population (see figure 11, inserted plot). For this group of typical decision makers, the theoretical probability to predict more than 85% of the answers is 2.8%. Similarly to the subjects whose distributions are shown in figure 10, for most decision makers in the experiment, we found prospect probabilities for which it was very unlikely to predict more than 85% of the answers. Figure 11 presents the frequencies, among all 142 subjects, of the probability of the theoretical predicted fraction of choices P^i to be larger than 85%. From this figure, we can extract the following representative statistics: for 56% of the population (80 subjects), the theoretical probability to predict correctly more than 85% of the choices (i.e. $P^i > 85\%$) is less than 5%; for 42% of the decision makers (60 subjects), the probability of $P^i > 85\%$ is less than 1%; for 28% (40 subjects), it is less than 0.1%. Consequently, even if the decision maker’s preferences are stable and if the estimated probabilities are very accurate, the probabilistic nature of the approach does not allow one to improve the choice predictions beyond its theoretical limit (which remains randomly distributed).

4.2 Distribution of predicted fractions at the aggregate level

Since only one predicted fraction at time 2 is observed for each subject, it is not possible to verify at the individual level whether the predicted fraction P^i of choices really follows the Poisson binomial distribution described in the previous subsection. However, assuming that the subjects belong to an homogeneous population[4], it is possible to approximate the distribution of the predicted fraction throughout the population, and to compare it to the histogram of the 142 observed predicted fractions at time 2.

For this purpose, we now consider that the Poisson binomial distribution of the fraction P^i of choices predicted correctly for subject i can be approached with the classical binomial distribution $\mathcal{B}\left(p^i, N\right)$, where p^i is defined by (see

---

[4] As discussed in sections 2.2.3 and 3.5, this assumption is not perfect but is useful as a first-order approximation.

figure 12, left panel):

$$p^i = \frac{1}{91}\left\lfloor \sum_{j=1}^{91} p_j^i \right\rfloor \in \left\{\frac{45}{91}, \frac{47}{91} \dots \frac{91}{91}\right\} \tag{20}$$

Moreover, we assume that the probability to pick a subject such that $P^i \sim \mathcal{B}(k/N, N)$, with $k \in \{45, \dots 91\}$, is equal to the frequency with which $p^i = k/N$ (figure 12, right panel). For each subject, this observed average prospect probability p^i of the most likely choice ($p_j^i > 0.5$) among 91 pairs of lotteries is estimated at time 1 with the QDT model. These approximations provide accurate representations of the results.

The theoretical distribution of the predicted fraction of choices throughout the population (142 subjects) is estimated by approximating binomial distributions, with success probabilities in the interval $(0.5; 1]$, which are then weighted by the observed frequencies of the average prospect probabilities p^i (see figure 13). Assuming that the prospect probabilities estimated at time 1 are accurate and stable in time and can thus be used at time 2 (to perform an "out-of-sample" prediction), the obtained theoretical distribution of the predicted fraction of choices in the population is given by the black solid line in figure 14. The red histogram corresponds to the predicted fractions observed at time 2. The approximated theoretical distribution for the predicted fraction appears to be close to the experimental one. In particular, both are skewed to the left: this suggests that bad predictions at the individual level may follow inevitably from the probabilistic nature of the choice.

Table 6: Estimated and experimental moments of the predicted fractions throughout the population.

	Approximated distribution	Experimental distribution
Mean	0.75	0.74
Standard deviation	-0.09	-0.09
Skewness	-0.3	-0.8

Performing the Kolmogorov-Smirnov test to compare the theoretical and observed distributions of predicted fraction shown in figure 14, we fail to reject at the 5% significance level the null hypothesis that the experimental distribution of the predicted fraction is generated by the theoretical one: the p-value is 0.254, and the value of the test statistic is 0.08 (corresponding to the maximum distance shown by the arrow in figure 15.

Table 6 and figure 15 compare the estimated cumulative distribution function (CDF) of the predicted fraction and the experimental one. Though the Kolmogorov-Smirnov test fails to reject the null hypothesis as just mentioned, this figure highlights a difference between the two CDF: the theoretical CDF seems to almost dominate stochastically the experimental one, i.e., the predicted fractions are less good than expected. The reason may be that the model is slightly overfitting. In other words, if a subject picks lottery A with probability $p_A > 0.5$, and actually chooses A at time 1, then maximising the likelihood might lead to overestimating p_A, thereby overestimating the "probability of success" when making the prediction that the subject will choose A at time 2.

5 Conclusion

We have analysed an experimental data set comprising 91 choices between two lotteries (two"propects") presented in random pairs made by 142 subjects repeated at two separated times. We have proposed an original quantification of the choice reversals occurring between the two repetitions, which provides a novel support for an intrinsic probabilistic approach to decision making. This has motivated us to test for the quantitative performance of a certain parameterisation of quantum decision theory (QDT). As predicted by QDT, we found that the stability of the prospect probabilities at the aggregate level is accompanied by variability of individual choices. In particular, for the majority of the pairs of lotteries, a significant proportion of subjects shifted their choices between two iterations of the experiment. The observed frequency of shifts was found in remarkable agreement with the prediction of a probabilistic choice theory, given the fact that it has no adjustable parameters and the comparison is therefore not a fit. Introducing heterogeneity between decision makers through a differentiation of the population into two similar sized groups in terms of over-confident and contrarian decision makers, we found an excellent quantitative description of the observed frequency of choice shifts.

Presenting a synthetic formulation of the main ingredients of QDT in the Appendix, we provided a novel constraint of the attraction factor q for a set of two prospects: $|q| \leq \min(f, 1-f)$, where f is the utility factor. The new bounds for q are more restrictive than previously considered $\{-1; 1\}$, and are sufficient for insuring the general condition $f + q \in [0, 1]$.

This study pioneered a parametric analytical formulation of QDT, integrating elements of (a) a stochastic version of Cumulative prospect theory (logit-CPT) for the utility factor f, and (b) constant absolute risk aversion (CARA) for the attraction factor. In essence, this approach allows one to separate risk aversion to extremely big losses, and transfer it into the QDT attraction factor. As a consequence, comparing with the benchmark, i.e., the logit-CPT implementation of Murphy and ten Brincke (2017), the loss aversion parameter λ was found to be smaller for the QDT model, while the values of the other parameters (α, δ, γ) remained close to those found for the logit-CPT model. The proposed QDT model improves the results of the logit-CPT model at both individual and aggregate levels, and for all criteria (explanatory power, predictive power, goodness of fit). The accentuation of the aversion to extreme losses embodied by the QDT attraction factor allowed us to noticeably improve the prediction of choices for the pairs of lotteries involving large losses. Thus, QDT transcends current theories of decision making under risk because it does not assume that risk aversion is a stable trait of a person. In contrast, it assumes that the overall risk aversion of a person is fixed (as assumed by the utility factor), but it allows for significant variability as a function of conditions embodied by the state of mind of the decision maker. These fluctuations are incorporated in the attraction factor.

At the same time, for most pairs of lotteries, the improvement was rather small. This is however hardly unique as there seems to exist a saturation of the average predicted fraction of choices at about 73-74% within the investigated probabilistic frameworks. We showed that this hard "barrier" is an intrinsic consequence of stochasticity in decision making, thus providing additional support for an inherent probabilistic component of choice making.

To quantify the limits of predictability, we proposed the Poisson binomial distribution as the theoretical distribution of the individual predicted fraction of correct choices. Then, for most decision makers in the experiment, we found the prospect probabilities for which it was very unlikely to predict more than 85% of the answers. Since only one predicted fraction is observed for each subject during the experiment, this theoretical distribution cannot be verified at the individual level. Thus, the distribution of the predicted fraction over the whole population was approximated with binomial distributions, and was found to be close to the experimental distribution of the predicted fractions over the 142 subjects. The Kolmogorov-Smirnov test did not reject the null hypothesis that the experimental distribution of the predicted fraction is the same as the theoretical one. However, the experimental fractions are slightly worse than expected, which may indicate that some subjects changed their state of mind, thus being less predictable that we assumed. Both distributions are skewed to the left, suggesting an intrinsic difficulty in predicting stochastic individual choices. Finally, heterogeneity between subjects might also explain these slight discrepancies.

The simplicity of QDT lies in the decomposition (9) in which appears the novel attraction factor. To strengthen the evidence provided here, it would be useful to test different forms of the utility factor, as the use of the logit-CPT model may be a weakness of the test of QDT, in the sense that we have in fact presented a "joint" test of two parameterisations: (i) the use of the logit-CPT model for the utility factor and (ii) of a constant absolute risk aversion (CARA) function for the attraction factor. It is important to test other forms for the utility factor, such as regret theory that has less parameters and develop a similar horse-race between regret theory alone and regret theory with the attraction factor. Many other combinations should be explored.

Acknowledgements We are grateful to R.O. Murphy and R.H.W. ten Brincke for the provided experimental data and useful discussions. This work was partially supported by the Swiss National Foundation, under grant 105218_159461 for the project on "Quantum Decision Theory".

Appendix: Quantum decision theory (QDT)

Developed by Yukalov and Sornette in a series of articles (Yukalov and Sornette, 2008, 2009, 2010, 2015a), quantum decision theory (QDT) has recently been introduced as an alternative formulation to existing theories. It is based on two essential ideas: (i) an intrinsic probabilistic nature of decision making and (ii) a generalisation of probabilities using the mathematics of Hilbert spaces that naturally accounts for entanglement between choices.

Mathematical structure of QDT

Let us recall briefly the mathematical construction of quantum decision theory (which can be found in more details in (Yukalov and Sornette, 2010)).

- *Definitions: actions, prospects and state of mind*

Definition 1 (Action ring) The action ring $\mathcal{A} = \{A_n : n = 1, 2, \ldots, N\}$ is the set of intended actions, endowed with two binary operations:

- The reversible and associative addition.
- The non-distributive and non-commutative multiplication, which possesses a zero element called empty action.

The interpretation of the sum $A + B$ is that A or B is intended to occur. The product AB means that A and B will both occur. The zero element is the impossible action, so $AB = BA = 0$ means that the actions A and B cannot occur together: they are disjoint.

Definition 2 (Composite action and action modes) When an action A_n can be represented as an union (i.e. is the sum of several actions), it is referred to as composite. Otherwise it is simple.
The particular ways A_{jn} of realizing a composite action A_n are called the action modes and are disjoint simple elements:

$$A_n = \bigcup_{j}^{M_n} A_{jn} \quad M_n > 1 \tag{21}$$

Definition 3 (Elementary prospects) An elementary prospect e_α is an intersection of separate action modes,

$$e_\alpha = \bigcap_n A_{\alpha n} \tag{22}$$

where the $A_{\alpha n}$ are action modes such that $e_\alpha e_\beta = 0$ if $\alpha \neq \beta$.

Definition 4 (Action prospect) A prospect π_n is an intersection of intended actions, each of which can be simple (represented by a single action mode) or composite

$$\pi_n = \bigcap_j A_{n_j} \tag{23}$$

To each action mode, we associate a mode state $|A_{jn}\rangle$ and its hermitian conjugate $\langle A_{jn}|$. Action modes are assumed to be orthogonal and normalized to one, so that $\langle A_{jn} \mid A_{kn}\rangle = \delta_{jk}$. This allows us to define orthonornal basic states for the elementary prospects:

$$|e_\alpha\rangle = |A_{\alpha 1} \dots A_{\alpha N}\rangle \quad \text{and} \quad \langle e_\alpha \mid e_\beta\rangle = \prod_n \delta_{\alpha_n}\delta_{\beta_n} = \delta_{\alpha\beta} \tag{24}$$

Definition 5 (Mind space and prospect state) The mind space is the Hilbert space

$$\mathcal{M} = \text{Span}\{|e_\alpha\rangle\} \ . \tag{25}$$

For each prospect π_n, there corresponds a prospect state $|\pi_n\rangle \in \mathcal{M}$

$$|\pi_n\rangle = \sum_\alpha a_\alpha |e_\alpha\rangle \ . \tag{26}$$

Definition 6 (Strategic state of mind) The strategic state is a normalized fixed state of the mind space $\mathcal{M}$ describing a decision maker at a given time:

$$|\psi\rangle = \sum_\alpha c_\alpha |e_\alpha\rangle \ . \tag{27}$$

The strategic state characterizes a particular decision maker at a given time, it includes his/her personal attributes and is related to the information available to the decision maker.

- *Prospect probabilities*

In the context of quantum decision theory, the preferences of a decision maker depend on his/her state of mind and on the available prospects. Those preferences are expressed through prospect operators.

Definition 7 (Prospect operator) For each prospect π_n, we define the prospect operator

$$\hat{P}(\pi_n) = |\pi_n\rangle\langle\pi_n| \ . \tag{28}$$

By this definition, the prospect operator is self-adjoint. Its average over the state of mind defines the prospect probability $p(\pi_n)$:

$$p(\pi_n) = \left\langle \psi \mid \hat{P}(\pi_n) \mid \psi \right\rangle \ . \tag{29}$$

The decision maker is more likely to choose the prospect with the highest prospect probability. The probabilities should correspond to the frequency with which the prospect would be chosen if the choice could be made several times in a same state of mind.

By definition 5 and 6, we can distinguish two terms in the expression of $p(\pi_n)$: a utility factor $f(\pi_n)$ and an attraction factor $q(\pi_n)$:

$$p(\pi_n) = f(\pi_n) + q(\pi_n) \tag{30}$$

$$f(\pi_n) = \sum_{\alpha} | c_\alpha^* a_\alpha |^2 \tag{31}$$

$$q(\pi_n) = \sum_{\alpha \neq \beta} c_\alpha^* a_\alpha a_\beta^* c_\beta \tag{32}$$

Within the framework of quantum decision theory, the utility and attraction terms are subjected to additional constraints:

- $f(\pi_n) \in [0,1]$ and $\sum f(\pi_n) = 1$ (normalization of the utility factor) ,
- $q(\pi_n) \in [-1,1]$ and $\sum q(\pi_n) = 0$ (alternation property of the quantum factor) .

Novel constraint of the attraction factor for a set of two prospects

The QDT formulation for a set of two prospects is now presented, and a new constraint for the attraction factor q is derived.

In the case of the choice between two lotteries (prospects) A and B (see table 1), the constraints on f and q can be written simply:

$$\begin{cases} p_A = f_A + q_A \\ p_B = f_B + q_B \\ q_A = -q_B \\ f_A = 1 - f_B \end{cases} \tag{33}$$

The goal being to calibrate quantum decision theory to the decisions made on pairs of lotteries, it is important to make some additional assumptions on the prospects involved.

Thus, we suppose that the prospects corresponding to the pairs of lotteries presented in table 1 can be written as follows:

$$\begin{cases} |A\rangle = a_1|A1\rangle + a_2|A2\rangle \\ |B\rangle = b_1|B1\rangle + b_2|B2\rangle \end{cases} \tag{34}$$

where $|A1\rangle$, $|A2\rangle$, $|B1\rangle$ and $|B2\rangle$ are orthogonal action mode states (this decompsition might be linked to the coexistence of belief and disbelief as suggested in (Yukalov and Sornette, 2015a), but the precise content of these action mode states will not be specified here).

We write the state of mind as

$$|\psi\rangle = c_{A_1}|A1\rangle + c_{A_2}|A2\rangle + c_{B_1}|B1\rangle + c_{B_2}|B2\rangle \tag{35}$$

and we denote by f_{A_1} and f_{A_2} the following quantities

$$f_{A_1} = |c^*_{A_1}a_1|^2; \quad f_{A_2} = |c^*_{A_2}a_2|^2 \ . \tag{36}$$

Then, the utility factor f_A satisfies

$$f_A = f_{A_1} + f_{A_2} \ . \tag{37}$$

Moreover, according to equation (32), the attraction factor is such that

$$q_A = c^*_{A_1}a_1a^*_2c_{A_2} + c^*_{A_2}a_2a^*_1c_{A_1} = 2\,\mathrm{Re}\left(c^*_{A_1}a_1a^*_2c_{A_2}\right) \tag{38}$$

Consequently, we can introduce the *uncertainty angle* Δ^A (Yukalov and Sornette, 2009) such that

$$q_A = 2\sqrt{f_{A_1}f_{A_2}}\cos\left(\Delta^A\right) \tag{39}$$

Moreover, equations (36) and (37) imply that there exists some $x \in [0,1]$ such that

$$f_{A_1} = xf_A \text{ and } f_{A_2} = (1-x)\,f_A \tag{40}$$

So, for some $x \in [0,1]$, we have that

$$q_A = 2f_A\sqrt{x\,(1-x)}\cos\left(\Delta^A\right) \tag{41}$$

In particular, $|q_A| \leq f_A$, and the same reasoning for the lottery B gives $|q_B| \leq f_B = 1 - f_A$. Consequently, given that $|q_A| = |q_B|$, we obtain that

$$|q_A| \leq \min\left(f_A, 1 - f_A\right) \tag{42}$$

Therefore, assumption (34) leads to a novel constraint for the attraction factor of quantum decision theory for a set of two prospects, which is given by equation (42). The bounds for q_A are found to be more restrictive than $\{-1, 1\}$, and are sufficient to insure the general condition $f_A + q_A \in [0,1]$.

Analytical formulation for the calibration of QDT

Under the assumptions done in the previous subsection, the formulation of QDT for choices between two lotteries A and B should be such that:

- $f_A = 1 - f_B$ (normalization)
- $q_A = -q_B$ (alternation)
- $q_A = \min\left(f_A, f_B\right)\cos\left(\Delta^A\right)$ (uncertainty factor)

The two next subsections address the parametrisation chosen for the utility term f and the attraction term q.

Utility term and stochastic cumulative prospect theory

Since the f-factor should represent a normalized utility, it is a natural choice to make it correspond to a stochastic version of cumulative prospect theory (CPT). Prospect theory was introduced by Kahneman and Tversky (1979) and is now the most famous alternative to expected utility theory. Within this framework, the outcomes are transformed through a value function v, and the probabilities are modified through a non-additive weighting function w. Moreover the value function separates gains and losses, where the notions of gains and losses are defined with respect to a reference point, here assumed to be zero. Cumulative prospect theory (CPT) is a variation of prospect theory, in which the weighted probabilities for outcomes of same sign should sum up to 1 (Kahneman and Tversky, 1992).

With this CPT model, for the simple pairs of lotteries shown in table 1, a lottery A is valued by

$$\tilde{U}_A = \begin{cases} w\left(p_1^A\right) v\left(V_1^A\right) + \left(1 - w\left(p_1^A\right)\right) v\left(V_2^A\right) & \text{if } sign(V_1^A) = sign(V_2^A) \\ w\left(p_1^A\right) v\left(V_1^A\right) + w\left(p_2^A\right) v\left(V_2^A\right) & \text{if } sign(V_1^A) \neq sign(V_2^A) \end{cases} \tag{43}$$

where V_1^A, V_2^A have been ordered such that:

- $V_1^A \geq V_2^A$ if both are positive.
- $V_1^A \leq V_2^A$ if both are negative.

The value function v is chosen to be convex in the domain of losses and concave in the domain of gains. This properties reflect commonly observed behavioural patterns: risk aversion concerning gains, and risk seeking behaviour with respect to losses.
For probability weighting, different formulations tend to suggest an inverse-S shaped function, so that small probabilities are overweighted and large probabilities underweighted.

In the present article, the value function is a power function with the same exponent α in the gain and the loss domains with a kink at 0 quantified by the loss aversion coefficient λ:

$$v(x) = \begin{cases} x^{\alpha} & x \geq 0 \quad \alpha > 0 \\ -\lambda(-x)^{\alpha} & x < 0 \quad \lambda > 0 \end{cases} \tag{44}$$

For probability weighting, a function known as the Prelec II weighting function was chosen (Prelec, 1998). It includes two parameters: δ controls the general elevation of the curve, and γ controls its curvature.

$$w(p) = \exp\left(-\delta\left(-\ln(p)\right)^{\gamma}\right) \ , \qquad \delta > 0 \quad \gamma > 0 \ . \tag{45}$$

Given the individual preferences and characteristics of a decision maker, CPT aims at predicting the choice of a decision maker, assuming it to be deterministic in the sense that the decision corresponds to the option that maximises the outcomes values weighted by subjective probabilities. In order to account for the ubiquitous observation of choice stochasticity, CPT can be combined with a probabilistic choice function. In particular, the stochastic version of CPT incorporates the probabilistic deviation of a decision maker from the option that maximises the choice criterion with respect to alternative options. Stott (2006) investigated several possible combinations and confirmed that a power value function (44) combined with a Prelec II weighting function (equation 45) is a good choice[5], when it is supplemented by a logit choice function (referred to as logit-CPT). The probability predicted by stochastic CPT of picking an option A over B is assumed to coincide with the f-factor of QDT and is given by

$$f_A = \frac{1}{1 + e^{\varphi\left(\tilde{U}_B - \tilde{U}_A\right)}} \ , \tag{46}$$

where φ is a sensitivity parameter.

According to this formulation of the f-factor given by stochastic cumulative prospect theory, the utility factor of QDT can be characterized by five parameters: two for the value function (α, λ), two for the weighting function (γ, δ) and one for the choice function (φ). This formulation of the value and probability weighting functions as well as the stochastic component is identical to that used by Murphy and ten Brincke (2017) and allows for a straightforward comparison of their results with our calibration of QDT. Indeed, when the attraction factor q is vanishing, QDT then reduced to stochastic CPT.

Attraction factor

As for the attraction factor, we have

$$q_A = \min\left(f_A, f_B\right) \cos\left(\Delta^A\right) \qquad \text{with } q_A + q_B = 0 \ . \tag{47}$$

The main issue is then to find a good parametrisation of the uncertainty factor $\cos\left(\Delta^A\right)$ that adds useful information, without adding too many parameters. In the current study, we replaced the cosine by

$$\cos\left(\Delta^A\right) = \tanh\left(a\left(U_A - U_B\right)\right) \ , \tag{48}$$

where U_A and U_B are utilities associated with the lotteries A and B that need to be specified, and a is either an additional parameter or a pre-defined constant. Thus,

$$q_A = \min\left(f_A, f_B\right) \tanh\left(a\left(U_A - U_B\right)\right) \tag{49}$$

---

[5] Several formulations of stochastic CPT were ranked in (Stott, 2006). The logit-power-Prelec II combination appeared to offer a good tradeoff between quality of fit and number of parameters.

This formulation satisfies automatically the alternating condition $q_A = -q_B$. To be specific, we assume that U is the constant absolute risk aversion (CARA) function for an initial wealth of 100 corresponding to the amount given to the subjects at the beginning of the experiment:

$$U(V) = 1 - e^{-\eta(100+V)} \tag{50}$$

With this formulation, q_A tends to be negative when the lottery A involves big losses and is compared to a lottery B with more moderate losses.

Bibliography

N.I. Al-Najjar and J. Weinstein. The ambiguity aversion literature: A critical assessment. *Economics and Philosophy*, 25(3):249–284, 2009.

M. Allais. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, 21(4):503–46, 1953.

M. Ashtiani and M.A. Azgomi. A survey of quantum-like approaches to decision making and cognition. *Math. Soc. Sci.*, 75:49–50, 2015.

F. Bagarello. *Quantum Dynamics for Classical Systems.* Wiley, Hoboken, 2013.

T. Ballinger and N. Wilcox. Decisions, error and heterogeneity. *Economic Journal*, 107:1090–1105, 1997.

D. Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):23–36 (translated by Dr. L. Sommer in 1954), 1738.

M. H. Birnbaum. Evidence against prospect theories in gambles with positive, negative, and mixed consequences. *Journal of Economic Psychology*, 27: 737–761, 2006.

M.H. Birnbaum. The nonadditivity of personality impressions. *Journal of Experimental Psychology*, 102(3):543, 1974.

M.H. Birnbaum. New paradoxes of risky decision making. *Psychological Review*, 115(2):463–501, 2008.

M.H. Birnbaum and A. Chavez. Test of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes*, 71(2):161–194, 1997.

P. Blavatskyy and G. Pogrebna. Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics*, 25:963–986, 2010.

P.R. Blavatskyy. A stochastic expected utility theory. IEW - Working Papers 231, Institute for Empirical Research in Economics - University of Zurich, 2005.

H.D. Block and J. Marschak. Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, H. Hoeffding, W. Madow, and H. Mann, editors, *Contributions to Probability and Statistics*, pages 97–132. Stanford University Press, Stanford, 1960.

S. Brown and A.J. Heathcote. The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology*, 57:153–178, 2008.

J.R. Busemeyer and P. Bruza. *Quantum Models of Cognition and Decision.* Cambridge University, Cambridge, 2012.

J.R. Busemeyer and J.T. Townsend. Decision field theory: A dynamic cognition approach to decision making. *Psychological Review*, 100:432–459, 1993.

J.R. Busemeyer, Z. Wang, A. Khrennikov, and I. Basieva. Applying quantum principles to psychology. *Phys. Scripta*, T163:014007, 2014.

C.F. Camerer. An experimental test of several generalized utility theories. *Journal of Risk and Uncertainty*, 2(1):61–104, 1989.

C.F. Camerer, G. Loewenstein, and R. Rabin. *Advances in Behavioral Economics.* Princeton University Press, Princeton, 2003.

E. Carbone and J.D. Hey. A comparison of the estimates of EU and non-EU preferences functionals using data from pairwise choice and complete ranking experiments. *Geneva Papers on Risk Insurance Theory*, 20:111–133, 1995.

E. Carbone and J.D. Hey. Which error story is best? *Journal of Risk and Uncertainty*, 20(2):161–176, 2000.

A. Conte, J. Hey, and P. Moffatt. Mixture models of choice under risk. *Journal of Econometrics*, 162:79–88, 2011.

C.P. Davis-Stober. Analysis of multinomial models under inequality constraints: Applications to measurement theory. *Journal of Mathematical Psychology*, 53:1–13, 2009.

W. Edwards. The prediction of decisions among bets. *Journal of Experimental Psychology*, 50(3):201–14, 1955.

W. Edwards. Subjective probabilities inferred from decisions. *Psychological Review*, 69(2):109–35, 1962.

J.-Cl. Falmagne. A stochastic theory for the emergence and the evolution of preference, structures. *Math. Social Sci.*, 31:63–84, 1996.

J.-Cl. Falmagne. Stochastic token theory. *J. Math. Psych.*, 41(2):129–143, 1997.

J.-Cl. Falmagne and S. Ovchinnikov. Media theory. *Discrete Appl. Math.*, 121:103–118, 2002.

S. Farrell and C.J.H. Ludwig. Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, 15 (6):1209–1217, 2008.

E. Fehr and U. Fischbacher. The nature of human altruism. *Nature*, 425: 785–791, 2003.

E. Fehr and S. Gächter. Cooperation and punishment in public good experiments. *The American Economic Review*, 90(4):980–994, 2000a.

E. Fehr and S. Gächter. Altruistic punishment in humans. *Nature*, 415:137–140, 2000b.

M. Fernández and S. Williams. Closed-form expression for the poisson-binomial probability density function. *IEEE Tansactions on Aerospace and Electronic Systems*, 46(2), 2010.

P.C. Fishburn. Stochastic utility. In S. Barbera, P.J. Hammond, and C. Seidl, editors, *Handbook of Utility*, pages 275–319. Kluwer Academic Publishers, Dordrecht, 2001.

B.U. Forstmann, R. Ratcliff, and E.-J. Wagenmakers. Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, 67:641–666, 2016.

S. Gaechter, E.J. Johnson, and A. Hermann. Individual-level loss aversion in risky and riskless choice. IZA Discussion Paper No. 2961, University of Nottingham, 2007.

I. Gilboa. Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics*, 16:65–88, 1987.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.

F. Glover. A user's guide to tabu search. *Annals of Operational Research*, 41 (3):3–28, 1993.

C. Gourieroux and A. Monfort. Testing non-nested hypotheses. In R.F. Engle and D.L. MacFadden, editors, *Handbook of Econometrics*, volume IV, pages 131–132. Elsevier, Oxford, 1994.

D.W. Harless and C.F. Camerer. The predictive value of generalized expected utility theories. *Econometrica*, 62:1251–1289, 1994.

E. Haven and A. Khrennikov. *Quantum Social Science*. Cambridge University, Cambridge, 2013.

J. Henrich. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, 53:3–35, 2004.

R. Hertwig and I. Erev. The description-experience gap in risky choice. *Trends in Cognitive Science*, 13(12):517–23, 2009.

M. Hetzer and D. Sornette. The co-evolution of fairness preferences and costly punishment. *PLoS ONE*, 8(3):e54308, pp.1–18, 2013a.

M. Hetzer and D. Sornette. An evolutionary model of cooperation, fairness and altruistic punishment in public good games. *PLoS ONE*, 8(11):e77041, pp. 1–13, 2013b.

J.D. Hey. Experimental investigations of errors in decision making under risk. *European Economic Review*, 39:633–640, 1995.

J.D. Hey. Does repetition improve consistency? *Experimental Economics*, 4: 5–54, 2001.

J.D. Hey. Why we should not be silent about noise. *Experimental Economics*, 8:325–345, 2005.

J.D. Hey and E. Carbone. Stochastic choice with deterministic preferences: An experimental investigation. *Economics Letters*, 47:161–167, 1995.

J.D. Hey and C. Orme. Investigating generalizations of expected utility theory using experimental data. *Econometrica*, 62(6):1291–1326, 1994.

D. Heyer and R. Mausfeld. On errors, probabilistic measurement and boolean valued logic. *Methodica*, 1:113–138, 1987.

D. Heyer and R. Niederee. Elements of a model-theoretic framework for probabilistic measurement. In E.E. Roskam, editor, *Mathematical Psychology in Progress*, pages 99–112. Springer, Berlin, 1989.

D. Heyer and R. Niederee. Generalizing the concept of binary choice systems induced by rankings: One way of probabilizing deterministic measurement structures. *Mathematical Social Sciences*, 23:31–44, 1992.

C. A. Holt and S.K. Laury. Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655, 2002.

A.I. Houston and J.M. McNamara. *Models of Adaptive Behaviour.* Cambridge University Press, Cambridge, 1999.

G.J. Iverson and J.-C. Falmagne. Statistical issues in measurement. *Mathematical Social Sciences*, 10:131–153, 1985.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

D. Kahneman and A. Tversky. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

D. Kahneman and A. Tversky. *Choices, Values, and Frames.* Cambridge University Press, Cambridge, 2000.

M.G. Kendall. On the reconciliation of theories of probability. *Biometrika*, 36:101–116, 1949.

A. Khrennikov. *Ubiquitous Quantum Structure.* Springer, Berlin, 2010.

A.N. Kolmogorov. *Foundations of the Theory of Probability.* English translation by Nathan Morrison, Chelsea, New York, 1956.

S. Lichtenstein and P. Slovic. Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89(1):46–55, 1971.

H.R. Lindman. Inconsistent preferences among gambles. *Journal of Experimental Psychology*, 89:390–97, 1971.

G. Loomes. Modelling the stochastic component of behaviour in expriments: Some issues for the interpretation of the data. *Experimental Economics*, 8: 301–323, 2005.

G. Loomes. Variability, noise, and error in decision making under risk. In G. Keren and G. Wu, editors, *The Wiley Blackwell Handbook of Judgment and Decision Making*. John Wiley & Sons, 2015.

G. Loomes and G. Pogrebna. Testing for independence while allowing for probabilistic choice. *J. Risk Uncertain*, 49:189–211, 2014.

G. Loomes and G. Pogrebna. Do preference reversals disappear when we allow for probabilistic choice? *Management Science*, 63(1):166–184, 2017.

G. Loomes and R. Sugden. Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92(368):805–824, 1982.

G. Loomes and R. Sugden. A rationale for preference reversal. *American Economic Review*, 73:428–32, 1983.

G. Loomes and R. Sugden. Some implications of a more general form of regret theory. *Journal of Economic Theory*, 41(2):270–287, 1987.

G. Loomes and R. Sugden. Incorporating a stochastic element into decision theories. *European Economic Review*, 39:641–648, 1995.

G. Loomes and R. Sugden. Testing different stochastic specifications of risky choice. *Economica*, 65:581–598, 1998.

G. Loomes, P. G. Moffatt, and R. Sugden. A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty*, 24:103–130, 2002.

R.D. Luce. *Individual Choice Behavior*. Wiley, New York, 1959.

R.D. Luce. Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology*, 46:1–26, 1994.

R.D. Luce and L. Narens. Fifteen problems concerning the representational theory of measurement. In P. Humphreys, editor, *Patrick Suppes: Scientific Philosopher*, volume 2, pages 219–245. Kluwer Academic Publishers, Netherlands, 1994.

R.D. Luce and P. Suppes. Preference, utility, and subjective probability. In R.D. Luce, R.R. Bush, and E. Galanter, editors, *Handbook of Mathematical Psychology*, volume 3, pages 249–410. Wiley, New York, 1965.

M.J. Machina. Non-expected utility theory. In S.N. Durlauf and L.E. Blume, editors, *The New Palgrave Dictionary of Economics*. Macmillan, Basingstoke and New York, 2008.

A.A.J. Marley. Some probabilistic models of simple choice and ranking. *Journal of Mathematical Psychology*, 5:311–332, 1968.

A.A.J. Marley. A random utility family that includes many of the "classical" models and has closed form choice probabilities and choice reaction times. *British Journal of Mathematical & Statistical Psychology*, 42(1):13–36, 1989a.

A.A.J. Marley. A random utility family that includes many of the "classical" models and has closed form choice probabilities and choice reaction times: Addendum. *British Journal of Mathematical & Statistical Psychology*, 42 (1):280, 1989b.

A.A.J. Marley. A selective review of recent characterizations of stochastic choice models using distribution and functional equation techniques. *Mathematical Social Sciences*, 23(1):5–29, 1992.

J. Marschak. Binary-choice constraints and random utility indicators. In K.J. Arrow, S. Karlin, and P. Suppes, editors, *Proceedings of the first Stanford symposium on mathematical methods in the social sciences, 1959*, pages 312–329. Stanford University Press, Stanford, 1960.

M. Mäs and H.H. Nax. A behavioral study of "noise" in coordination games. *Journal of Economic Theory*, 162:195–208, 2016.

D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, 1974.

D. McFadden and K. Train. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15:447–470, 2000.

J.M. McNamara, P.C. Trimmer, and A.I. Houston. Natural selection can favour 'irrational' behaviour. *Biology Letters*, 10(1), 2014.

F. Mosteller and P. Nogee. An experimental measurement of utility. *Journal of Political Economy*, 59(5):371–404, 1951.

R.O. Murphy and R.H.W. ten Brincke. Hierarchical maximum likelihood parameter estimation for cumulative prospect theory: Improving the reliability of individual risk parameter estimates. *Management Science*, Published online in Articles in Advance: http://dx.doi.org/10.1287/mnsc.2016.2591, 2017.

J. Myung, G. Karabatsos, and G. Iverson. A bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, 49:205–225, 2005.

R. Niederee and D. Heyer. Generalized random utility models and the representational theory of measurement: A conceptual link. In A. A.J. Marley, editor, *Choice, Decision and Measurement: Essays in Honor of R. Duncan Luce*, pages 155–189. Lawrence Erlbaum, Mahwah, 1997.

H. Nilsson, J. Rieskamp, and E.-J. Wagenmakers. Hierarchical bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55:84–93, 2011.

T.J. Pleskac. Decision and choice: Luce's choice axiom. Technical report, Department of Psychology, Michigan State University, 2012.

D. Prelec. The probability weighting function. *Econometrica*, 66(3):497–527, 1998.

J. Quiggin. A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343, 1982.

J. Quiggin. *Generalized Expected Utility Theory. The Rank-Dependent Model.* Kluwer Academic Publishers, Boston, 1993.

M. Rabin. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68:1281–1293, 2000.

M. Regenwetter. Random utility representations of finite many relations. *Journal of Mathematical Psychology*, 40:219–234, 1996.

M. Regenwetter and C.P. Davis-Stober. Behavioral variability of choices versus structural inconsistency of preferences. *Psychological Review*, 119(2):408–416, 2012.

M. Regenwetter and A.A.J. Marley. Random relations, random utilities, and random functions. *Journal of Mathematical Psychology*, 45:864–912, 2001.

M. Regenwetter, J. Dana, and C.P. Davis-Stober. Testing transitivity of preferences on two-alternative forced choice data. *Frontiers in Psychology*, 1: 148, 1–15, 2010.

M. Regenwetter, J. Dana, and C.P. Davis-Stober. Transitivity of preferences. *Psychological Review*, 118(1):42–56, 2011.

M. Regenwetter, C.P. Davis-Stober, S.H. Lim, Y. Guo, A. Popova, C. Zwilling, Y.-C. Cha, and W. Messner. Qtest: quantitative testing of theories of binary choice. *Decision*, 1(1):2–34, 2014.

J. Rieskamp. The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6):1446–1465, 2008.

J. Rieskamp and P.E. Otto. Ssl: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135:207–236, 2006.

J. Rieskamp, J.R. Busemeyer, and B.A. Mellers. Extending the bounds of rationality: evidence and theories of preferential choice. *Journal of Economic Literature*, 44(3):631–661, 2006.

Z. Safra and U. Segal. Calibration results for non-expected utility theories. *Econometrica*, 76:1143–1166, 2008.

L.J. Savage. *The Foundations of Statistics.* Wiley, New York, 1954.

B. Scheibehenne and T. Pachur. Using bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review*, 22(2):391–407, 2015.

D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–87, 1989.

M. Schulte-Mecklenbeck, T. Pachur, R.O. Murphy, and R. Hertwig. Prospect theory tracks selective allocation of attention. Working paper, Max Planck Institute, Berlin, 2016.

D. Sornette. Physics and financial economics (1776-2014): Puzzles, Ising and agent-based models. *Rep. Prog. Phys.*, 77:062001, 2014.

D. Sornette. Decisions are deterministic. In B.S. Frey and D. Iselin, editors, *Economic Ideas You Should Forget*, pages 131–132. Springer, 2017.

C. Starmer and R. Sugden. Probability and juxtaposition effects: An experimental investigation of the common ratio effect. *Journal of Risk and Uncertainty*, 2(2):159–178, 1989.

N. Stewart, N. Chater, and G.D.A. Brown. Decision by sampling. *Cognitive Psychology*, 53:1–26, 2006.

H.P. Stott. Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, 32(2):101–130, 2006.

L.L. Thurstone. A law of comparative judgment. *Psychology Review*, 34: 273–286, 1927.

A. Tversky. Intransitivity of preferences. *Psychological Review*, 76(1):31–48, 1969.

A. Tversky and D. Kahneman. The framing of decisions and the dsychology of choice. *Science, New Series*, 211(4481):453–58, 1974.

J. von Neumann. *Mathematical Foundations of Quantum Mechanics.* Princeton University Press, Princeton, 1955.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior (2nd ed.).* Princeton University Press, Princeton, 1947.

P. Wakker. Formalizing reference dependence and initial wealth in Rabin's calibration theorem. *Working paper, Erasmus University*, 2005.

N.T. Wilcox. Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In J.C. Cox and G.W. Harrison, editors, *Risk Aversion in Experiments*, volume 12 (Book Series: Research in Experimental Economics), pages 197–292. Emerald Group Publishing Limited, 2008.

S.S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9:60–62, 1938.

J.I.J. Yellott. The relationship between luce's choice axiom, thurstone's theory of comparative judgement, and the double exponential distribution. *Journal of Mathematical Psychology*, 15:109–144, 1977.

V.I. Yukalov and D. Sornette. Quantum decision theory as quantum theory of measurement. *Physics Letters*, A 372:6867–6871, 2008.

V.I. Yukalov and D. Sornette. Processing onformation in quantum decision theory. *Entropy*, 11:1073–1120, 2009.

V.I. Yukalov and D. Sornette. Mathematical structure of quantum decision theory. *Advances in Complex Systems*, 13:659–698, 2010.

V.I. Yukalov and D. Sornette. Decision theory with prospect interference and entanglement. *Theory and Decision*, 70:283–328, 2011.

V.I. Yukalov and D. Sornette. Conditions for quantum interference in cognitive sciences. *Topics in Cognitive Science*, 6:79–90, 2014.

V.I. Yukalov and D. Sornette. Positive operator-valued measures in quantum decision theory. *Lecture Notes in Computer Science*, 8951:146–161, 2015a.

V.I. Yukalov and D. Sornette. Quantum probability and quantum decision making. *Phil. Trans. R. Soc. A*, 374:20150100, 2–15, 2015b.

V.I. Yukalov and D. Sornette. Preference reversal in quantum decision theory. *Frontiers in Psychology*, 6:01538, 2015c.

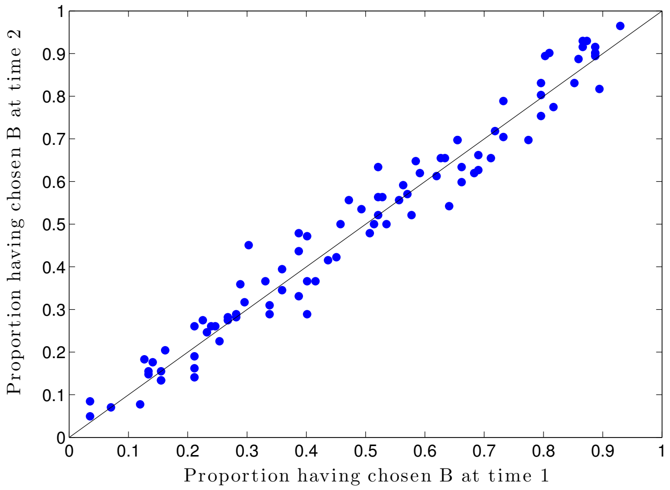

Fig. 1: Proportion of decision makers having chosen option B at time 2 as a function of the proportion of decision makers having chosen option B at time 1 (there are 91 points, one for each of the 91 presented pairs of lotteries)

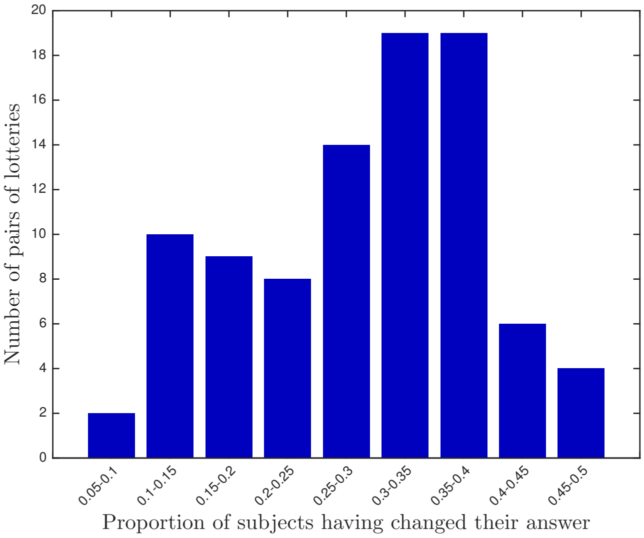

Fig. 2: Histogram over all 91 lottery pairs of the proportion of decision makers having changed their choice between times 1 and 2. Note that the ordinate values of the ten bins sum up to 91. For more than half of the considered lottery pairs (48 out of 91), more than 30% of the subjects shifted their preference from A to B or vice-versa, between times 1 and 2

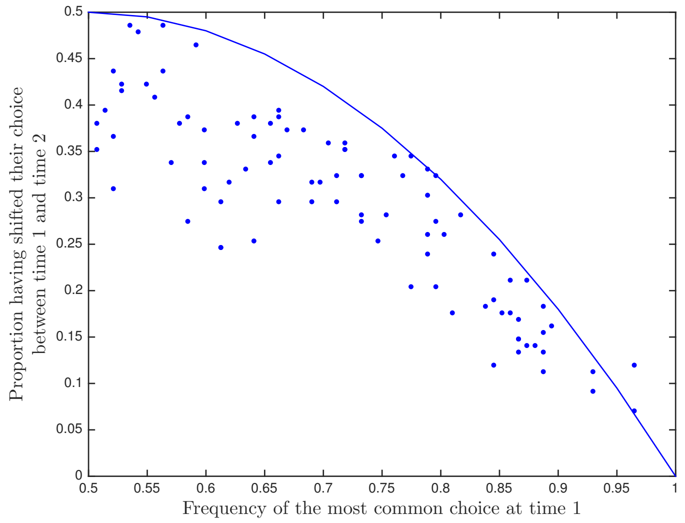

Fig. 3: Proportion of decision makers having shifted their choice between time 1 and time 2 as a function of the proportion choosing the most frequently chosen option at time 1 (there are 91 points, each one represents a pair of lotteries). The solid line represents the proportion of shifts predicted by expression (4) explained in the text. We stress that the solid line is not a "fit" as there are no adjustable parameters

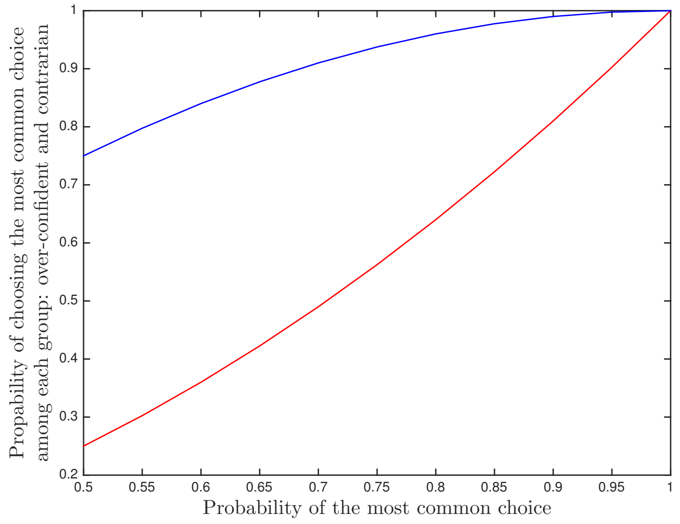

Fig. 4: Probabilities p_1 and p_2 given by equation (8) with which the most common choice is chosen for each of the postulated two groups of decision makers as a function of the average choice probability p aggregated over the whole population. In other words, the top (resp. bottom) curve shows the decision probability p_1 (resp. p_2) of the "over-confident" (resp. "contrarian") decision makers as a function of the frequency p of the most common choice

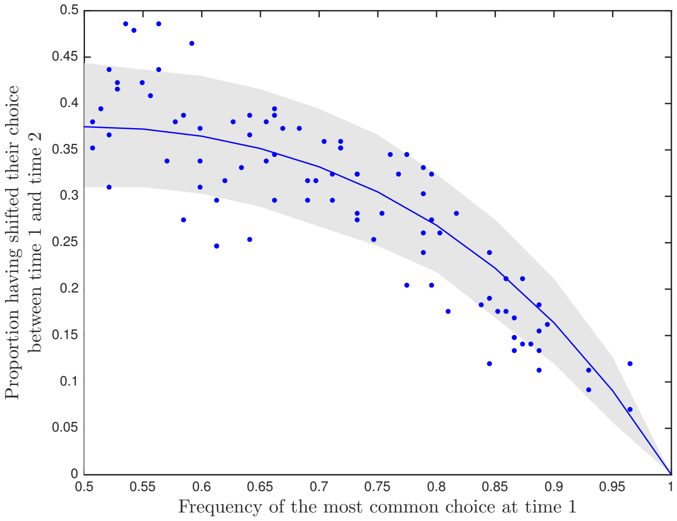

Fig. 5: Same data as figure 3, which is compared with the prediction (5) of an heterogeneous population of two groups of decision makers with p_1 and p_2 given by (8) with equal sizes $F = 1/2$ of the two groups. The shaded area represents the 5% and 95% quantiles, ie. the area where 90% of the shifts should fall according to Monte Carlo simulations using the above model with two groups (3000 simulations per pairs of lotteries)

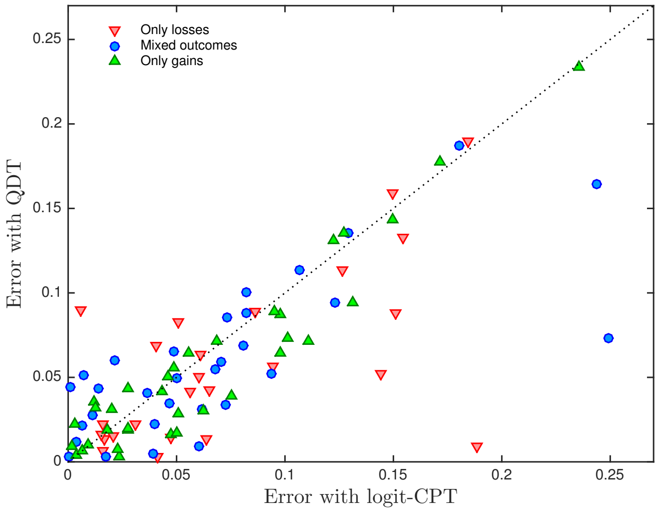

Fig. 6: Distances of the estimated choice frequencies (at the aggregate level) at time 2 to the choice frequencies observed at time 1. There are 91 points representing the 91 pairs of lotteries. Markers encode different types of gamble: only losses (red downward triangles), only gains (green upward triangles) and mixed-gambles (blue dots). The QDT model (y-axis) performs better than logit-CPT (x-axis) for gambles appearing in the lower triangle below the diagonal

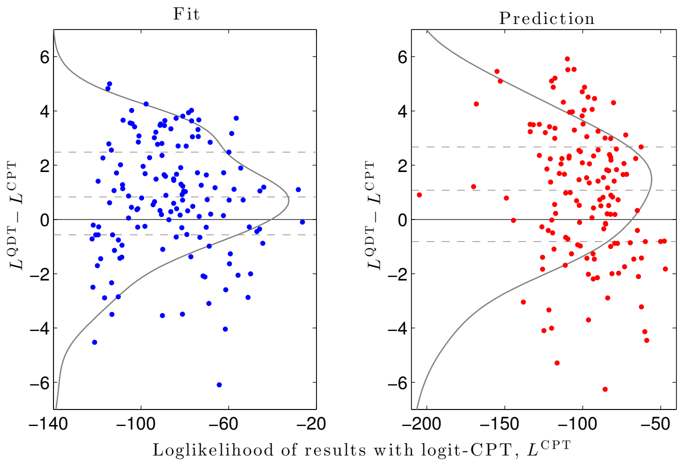

Fig. 7: Difference of log-likelihoods (y-axis, QDT minus logit-CPT) as a function of the log-likelihoods obtained with logit-CPT (x-axis). There are 142 points in each plot, each point represents a decision maker. QDT performs better for points with positive y-coordinate (i.e. when QDT leads to a larger log-likelihood). The left plot shows the results of the fit (time 1), and the right plot shows the results of the prediction (time 2). For both plots, the solid curves represent the kernel estimated density of the difference of log-likelihoods, and the dashed lines their median and quartiles

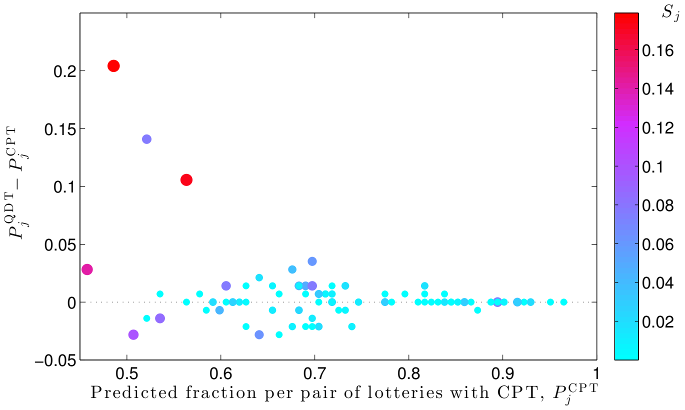

Fig. 8: P_j^{QDT} (resp. P_j^{CPT}) represents the fraction of choices correctly predicted for the pair j of lotteries with QDT (resp. logit-CPT). The difference between the two predicted fractions is plotted against the predicted fraction obtained with logit-CPT. For points in the upper part of the plot above the dotted line, QDT preforms better. The colors and sizes of markers encode the average intensity of the attraction factor among $N = 142$ subjects: $S_j = \frac{1}{N}\sum_{i=1}^{N} |q^i_{A_j}|$

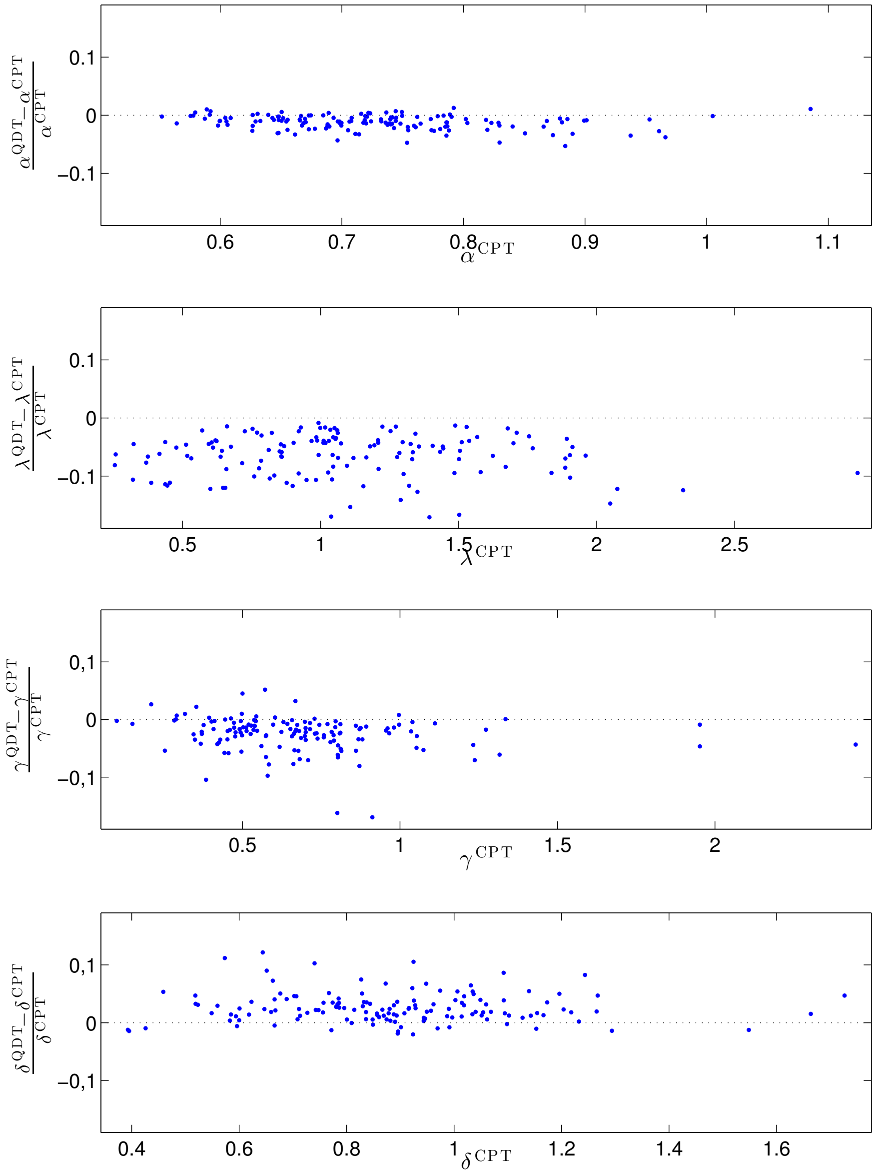

Fig. 9: Relative difference of parameters estimates (QDT minus logit-CPT) against estimates obtained with logit-CPT. Though for all parameters and most subjects the absolute relative difference are smaller than 10%, some trends are noticeable. In particular, in the presence of a quantum factor for extreme losses, the loss aversion λ tends to be smaller than with logit-CPT

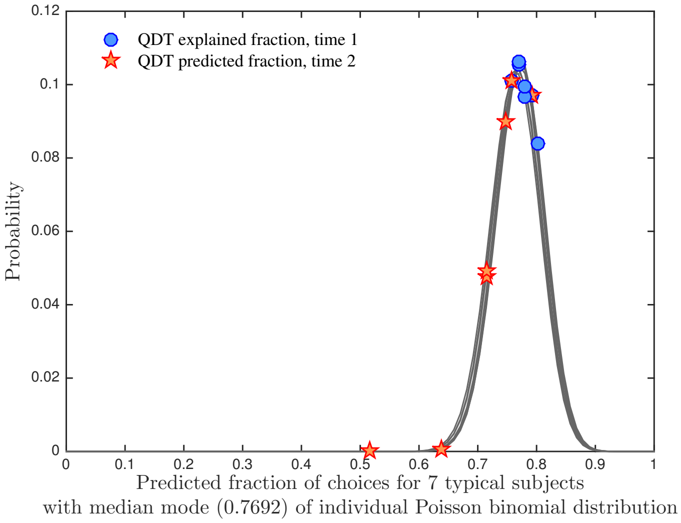

Fig. 10: Theoretical Poisson binomial distributions of the predicted fraction of choices, P^i, of a group of 7 typical (with their mode of P^i equal to 0.77, i.e. median value within the population: see figure 11, inset). For these theoretical distributions, individual prospect probabilities of the most likely choice ($p^i_j > 0.5$) for each of the 91 pairs of lotteries j's are estimated with the QDT model at time 1. The observed fractions of choices (i) explained at time 1 i.e. "in-sample" (blue circles) and (ii) predicted for time 2 i.e."out-of-sample (red pentagram) with the QDT model, are indicated on the plot. For this group of typical decision makers, the theoretical probability to predict more than 85% of the answers is 2.8%

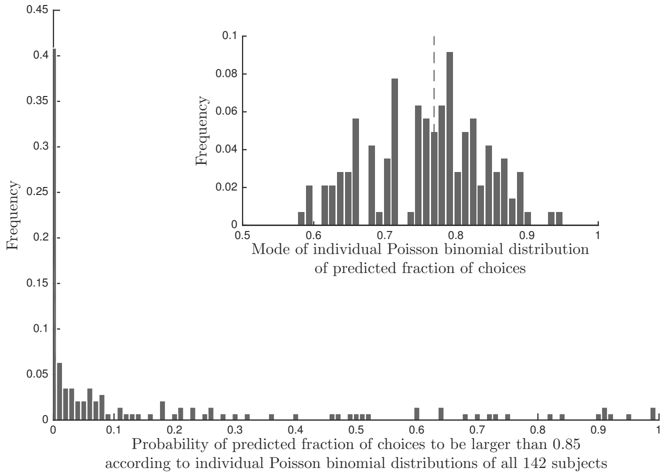

Fig. 11: Main plot: Frequencies, over all 142 subjects, of the probability of the theoretical predicted fraction of choices P^i to be larger than 85%. For 56% of the population (80 subjects), the theoretical probability to predict more than 85% of choices is less than 5%. **Inset**: Frequencies, over all 142 subjects, of the modes of theoretical individual Poisson binomial distributions of the predicted fraction of choices, with median value representing a "typical" decision maker indicated by dashed line

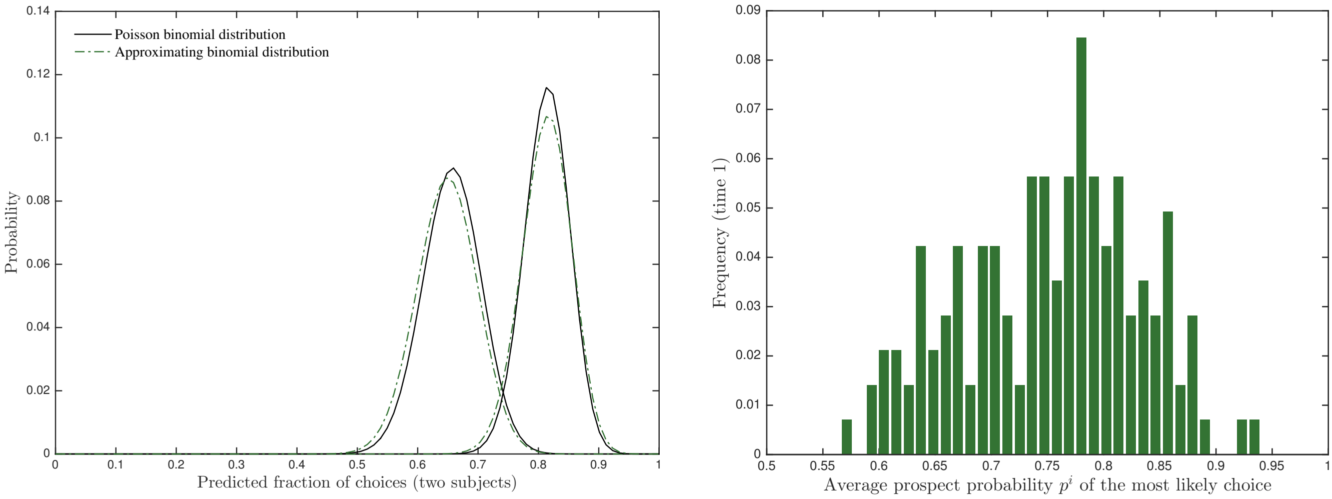

Fig. 12: Left: Theoretical Poisson binomial distributions (black solid line) of the predicted fractions of choices for two distinct subjects as described in subsection 4.1, and their approximating binomial distributions (green dash-dotted line). **Right**: Histogram, over 142 subjects, of their observed average prospect probabilities p^i of the most likely choice ($p^i_j > 0.5$) among 91 pairs of lotteries, estimated at time 1 with the QDT model (for each subject, p_i also corresponds to the mean of her theoretical predicted fraction of choices distributed according to the Poisson binomial law, and the approximating binomial law)

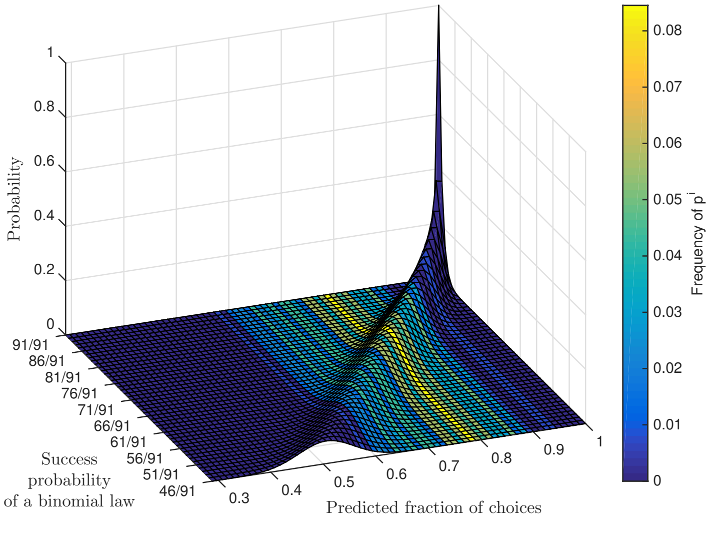

Fig. 13: Estimation of a theoretical distribution of the predicted fraction of choices throughout the population (142 subjects), which is obtained by combining the approximating binomial distributions, with success probabilities in the interval [46/91; 91/91], with weights determined by the observed frequencies of the average prospect probabilities p^i of the most likely choice at time 1 with QDT model (see figure 14)

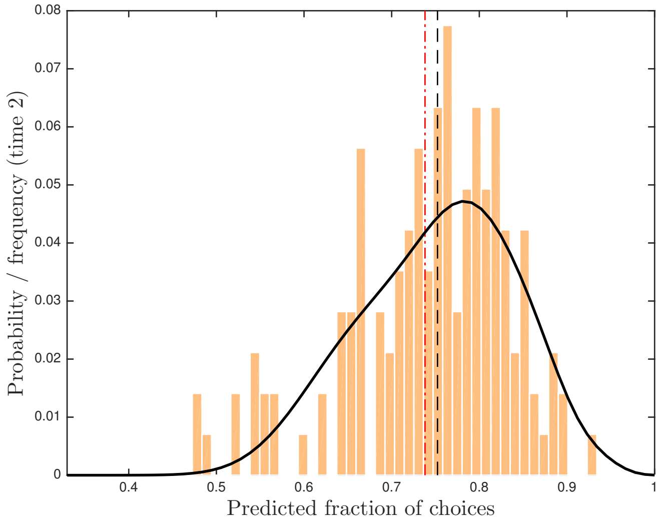

Fig. 14: Approximated theoretical distribution (black solid line) of the predicted fraction of choices throughout the population (142 subjects). The histogram represents the fractions of choices correctly predicted "out-of-sample" at time 2 with QDT estimated at time 1. Mean values are indicated by the black dashed line for the theoretical distribution and by the red dash-dotted line for the experimental distribution. These values are reported in table 6

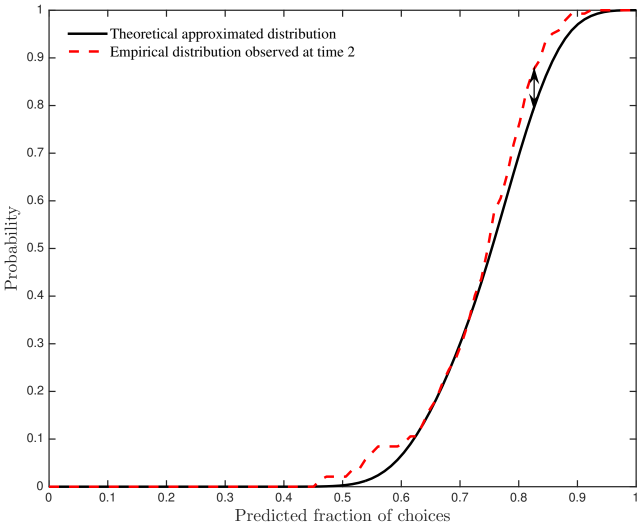

Fig. 15: Theoretical (black solid line) and experimental (red dashed line) cumulative distribution functions (CDF) of the predicted fractions of choices over the population (142 subjects). The arrow shows the maximum distance between the two curves (value of the test statistic for the Kolmogorov-Smirnov test equal to 0.08)

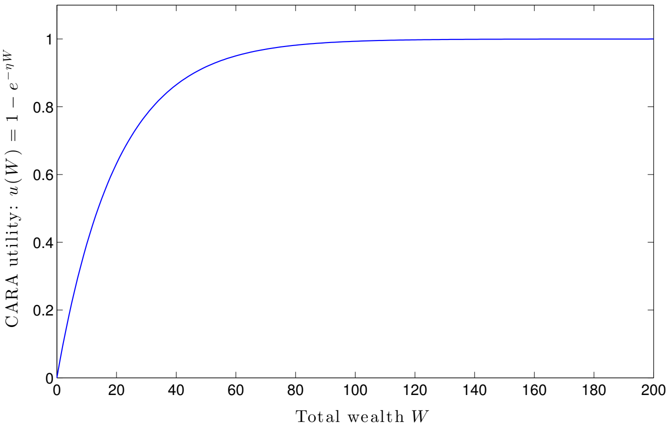

Fig. 16: CARA utility function for $\eta = 0.05$. The outcomes V defined in table 1 of the choices between pairs of lotteries in the experiments being between -100 and 100, and the initial given amount being 100, the total wealth W is considered to be in the interval $[0, 200]$. With this utility function and expression (49), we get that the attraction factor q is small except for pairs of lotteries involving big losses

The conjunction fallacy in quantum decision theory

Tatyana Kovalenko *[a] and Didier Sornette †[a,b]

[a]ETH Zurich, Department of Management, Technology and Economics, Scheuchzerstrasse
7, 8092 Zurich, Switzerland
[b]Swiss Finance Institute, c/o University of Geneva, 40 blvd. Du Pont d'Arve, CH 1211
Geneva 4, Switzerland

September 22, 2017

Abstract

The conjunction fallacy is a renowned violation of classical probability laws, which is persistently observed among decision makers. Within Quantum decision theory (QDT), such deviations are the manifestation of interference between decision modes of a given prospect. We propose a novel QDT interpretation of the conjunction fallacy, which cures some inconsistencies of a previous treatment, and incorporates the latest developments of QDT, in particular the representation of a decision-maker's state of mind with a statistical operator. Rather than focusing on the interference between choice options, our new interpretation identifies the origin of uncertainty and interference between decision modes to an entangled state of mind, whose structure determines the representation of prospects. On par with prospects, the state of mind can be a source of uncertainty and lead to interference effects, resulting in characteristic behavioral patterns.

We present the first in-depth QDT-based analysis of an empirical study (the touchstone experimental investigations of Shafir et al. (1990)), which enables a data-driven exploration of its underlying theoretical construct. We link typicality judgements to probability amplitudes of the decision modes in the state of mind, and quantify the level of uncertainty and the relative contributions of prospect's interfering modes to their probability judgement. This enables inferences about the key QDT interference "attraction" q-factor with respect to different types of prospects - compatible versus incompatible.

We propose a novel empirically motivated "QDT indeterminacy (or uncertainty) principle," as a fundamental limit of the precision with which certain sets of prospects can be simultaneously known (or assessed) by a decision maker, or elicited by an experimental procedure. For any type of prospects, we observe a general tendency for the q-factor to converge to the same negative range $q \in (-0.25, -0.15)$ in the presence of high uncertainty, which motivates the hypothesis of an universal "aversion" q. The "aversion" q is independent of the (un-)attractiveness of a prospect under more certain conditions, which is the main difference with the previously considered QDT "quarter law". The universal "aversion" q substantiates the previously proposed QDT uncertainty aversion principle and clarifies its domain of application. The universal "aversion" q provides a theoretical basis for modelling different risk attitudes, such as aversions to uncertainty, to risk or to losses.

Keywords. Quantum decision theory, conjunction fallacy, interference, indeterminacy (uncertainty) principle, universal aversion

---

*tkovalenko@ethz.ch (corresponding author)
†dsornette@ethz.ch

1 Introduction

The conjunction fallacy is a well-known behavioral pattern, when the probability for a conjunction category is judged larger than for its constituents. Although such probability judgement violates the axiomatic probability theory, it is nevertheless consistently observed among decision makers under different experimental setups (indirect, direct-subtle and direct-transparent tests). Several plausible explanations were proposed, such as fallacious representativeness and availability heuristics in the original study (Tversky and Kahneman, 1983), or "socially rational" semantic inferences about the meaning of 'probability' (Hertwig and Gigerenzer, 1999). Alternative approaches using the quantum formalism were also applied to tackle the problem. The quantum judgment model suggests that a choice is made by the sequential projection of a decision maker's belief state onto prospects subspaces, resulting in transition probabilities that are based on Lüder's rule (Busemeyer et al., 2011). However, the resultant 'question order effect' was criticized for its inability to model double conjunction fallacies. Moreover, empirical analyses confirmed the advantage of another approach to explain the conjunction fallacy, based on modelling 'states of conceptual entities' and 'emergence effects' (Aerts et al., 2017). Importantly, it can be shown rigorously that Lüder's rule for calculating the probability of consecutive measurements cannot be used for calculating quantum joint probabilities, in particular for non-commuting prospects. The rigorous derivation of quantum joint probabilities for arbitrary prospects, whether commuting or non-commuting, has been proposed within an 'emergence-type' Quantum decision theory (QDT) in (Yukalov and Sornette, 2013), which in addition indicates the differences between QDT and other quantum approaches.

However, the applications of QDT are currently limited by its complexity and the challenges in making it operational. Previous studies involved top-level aggregate experimental results, exploiting the most general QDT relation, $p = f + q$, where the probability p that a prospect is chosen is decomposed into the sum of two factors, the utility f and attraction q of that prospect. Preceding research either checked the agreement between data and that relation $p = f + q$ (Yukalov and Sornette, 2014), or sought to construct a sound parametrization of f and q, based on "classical" decision theories, such as Expected Value (Favre et al., 2016), Expected Utility Theory and (stochastic) Cumulative Prospect Theory (Vincent et al., 2017, Siffert et al., 2017).

According to (Yukalov and Sornette, 2009), the conjunction fallacy is explained by a non-attractiveness of a constituent category B (e.g. "bankteller"), by inclusion of an interference with a secondary feature A (e.g. "feminist"). However, the secondary feature is present only in a conjunction category AB (e.g. "feminist bankteller"). In indirect tests, as in (Shafir et al., 1990), a decision maker is exposed to only one of the judged categories at a time (a conjunction or its constituent). Thus, the negative interference with the secondary feature should not appear in deliberations concerning a single constituent category. In (Yukalov and Sornette, 2015), a general procedure to introduce uncertainty in decision making is proposed by incorporating a set $B \in \{B_1, B_2\}$, which represents the belief and disbelief of a decision maker about a task setup and a relevant criterion of choice. This approach may explain the occurrence of interference, but not their amplitude and the corresponding size of the effects. Thus, the conjunction fallacy remains unexplained within previous proposals using QDT.

Applying the quantum formalism to decision making and transforming its theoretical construct into an operational tool is challenging. A situation of choice, which includes a decision maker and choice options, should be characterized adequately in terms of the corresponding operators and state vectors, including: (a) a decision-maker's state of mind ρ (for pure states, $|\psi\rangle\langle\psi|$) and (b) prospects $|\pi_i>$. First, the choice of a basis for the representation of these vectors, i.e. elementary prospects $|e_j>$, is not trivial. For example, in a given experimental setup, a prospect can be postulated via certain action modes, however some of them can themselves turn out to be quantum superpositions. In addition, the state of mind can include interfering elementary prospects that are not originating from the prospects in consideration, but from other sources: (i) endogenous (a decision maker's own experience, beliefs, and so on) and (ii) exogenous (framing, environment, and so on). Second, revealing and estimating the coefficients (i.e. relative weights of decision modes) of the linear decomposition of the state of mind and of prospects on state vectors is a subtle task, as invasive elicitation methods can affect the state of mind of the decision maker.

In this paper, we focus on the experiment that was reported in (Shafir et al., 1990), and previously analyzed within QDT in (Yukalov and Sornette, 2009). We propose a novel QDT interpretation of the conjunction fallacy, which cures some inconsistencies of this previous treatment, and incorporates the latest developments of QDT, in particular the representation of a decision-maker's state of mind with a statistical operator (Yukalov and Sornette, 2015, 2016a). Our main contribution is to put the state of mind as the centre of the evaluation of the level of uncertainty that influence prospects' representations, resulting in interference effects. Our novel QDT interpretation of the conjunction fallacy is based on the following propositions:

1. Representativeness and availability heuristics are at the core of the conjunction fallacy (Tversky and Kahneman, 1983). The descriptions of subjects (instances I) and some of the associated categories (usually a secondary feature A in a conjunction AB) share common characteristics. After the exposition to an instance I, the state of mind of a decision maker is intentionally influenced (framed) by incepting into it specific elementary prospects, which then interfere with resembling choice options and modify the prospects' probabilities. Thus, the existence of an *interference between the state of mind and a prospect* is proposed to be the mechanism of the conjunction fallacy.

2. In order to calculate the probabilities of the prospects (explicitly, e.g. for the theoretical formulation of situations of choices, or implicitly, e.g. in real life situations), both prospects and state of mind should be represented with the same set of elementary prospects. Thus, in general, prospects and state of mind are *mutually dependent* in the granularity of their decompositions (i.e., they share the same basis of elementary prospects).

In practice, depending on the constituents (elementary prospects) of a state of mind, the modes of the intended actions, as formulated in a subsequent situation of choice, may require further decomposition, i.e. the presented modes in a given experiment may be in a *quantum superposition* themselves, or be a tensor product of several more specific (detailed) elementary prospects. In other words, the choice of the elementary prospects that form a basis *is context dependent*, and may differ from the explicitly presented formulation.

In summary, the present study differs from previous ones by first reinserting the state of mind at the core of the formalism, as it was initially defined in the axiomatic construction of Yukalov and Sornette (2009). This allows us to clarify the interference mechanism of QDT, based on the quantification of the decision maker's state of mind and of the considered prospects. This is achieved by linking typicality judgements and probability amplitudes of decision modes in the state of mind. We decompose the problem into several (extreme) cases, with minimum and maximum interference, as well as with singular or distributed weights of interfering modes in the state of mind. This allows us to estimate the level of uncertainty and the relative contributions of prospect's decision modes to their probability judgement. This level of granularity enables the analysis of broader and more nuanced datasets, when interference effects are less pronounced (for example, for prospects with compatible categories). It also opens the possibility of better characterising interdependencies and the relative importance of QDT elements, and finally makes the theory operational.

The organisation of the article is as follows. Section 2 summarises the analysed experimental setup and the main results, which will be compared with the prediction of our novel formulation in section 5. Section 3 recapitulates the previous approach to the conjunction fallacy developed by Yukalov and Sornette (2009) and dissects its weaknesses and inconsistencies. This opens up the road towards our new formulation, presented in section 4. Note that section 4 is self-contained and can be studied independently of section 3. Section 5 compares the detailed experimental results in the three main set-ups with the predictions of our new QDT formulation of the conjunction fallacy. Section 6 builds on previous sections and results to suggest a modification of the previous concept of an “attraction factor” q into a universal “aversion factor” q. Section 7 concludes.

2 Brief description of the analyzed experimental setup

The present study analyses empirical results reported in (Shafir et al., 1990). This section provides a brief description of the experimental setup and its main findings.

During the experiment, 110 decision makers were exposed to 14 instances I (description of a subject) followed by one of the four categories: (1) a compatible conjunction $AB(c)$; (2) its constituent $B(c)$; (3) an incompatible conjunction $AB(i)$; (4) its constituent $B(i)$. The compatibility (c) or incompatibility (i) type was attributed by the experimenters based on their qualitative evaluation of a number of shared properties between conjunction elements A and B.

An example of an instance I is the description of a subject: “Linda was a philosophy major. She is bright and concerned with issues of discrimination and social justice.”

The corresponding four categories are:

1. feminist teacher, $AB(c)$;
2. teacher, $B(c)$.
3. feminist bankteller, $AB(i)$;
4. bankteller, $B(i)$;

One group of participants (54 decision makers) was asked to make judgements about the typicality (typ_{exp}) of an instance I in each category. Another group (56 decision makers) provided their judgements about the probability (p_{exp}) that an instance I belonged to the corresponding category. Importantly, 14 instances were presented with each of the categories independently, i.e. an instance and one category at a time, and presented in mixed order. When sampling the four different categories for a given instance, these four pairs were presented randomly and were separated by pairs involving other instances.

Figure 1 illustrates the experimental setup (Shafir et al., 1990).

The conjunction effect was calculated by subtracting the typicality judgement of the instance with respect to a constituent category B from the corresponding typicality judgement with respect to the conjunction AB. The conjunction effect exists when this difference is positive. Similarly, the conjunction fallacy is qualified when the difference between the probability judgements for a conjunction AB and its constituent B is positive. Note that typicality and probability judgements were performed by two distinct groups of participants, which may lead to additional discrepancies between an intended theoretical interpretation (section 4) and the empirical values of prospects' probabilities.

The experimental results from (Shafir et al., 1990) are reproduced in appendix A.1, and their aggregate statistic is provided in table 1. Averaged among participants, the conjunction effect

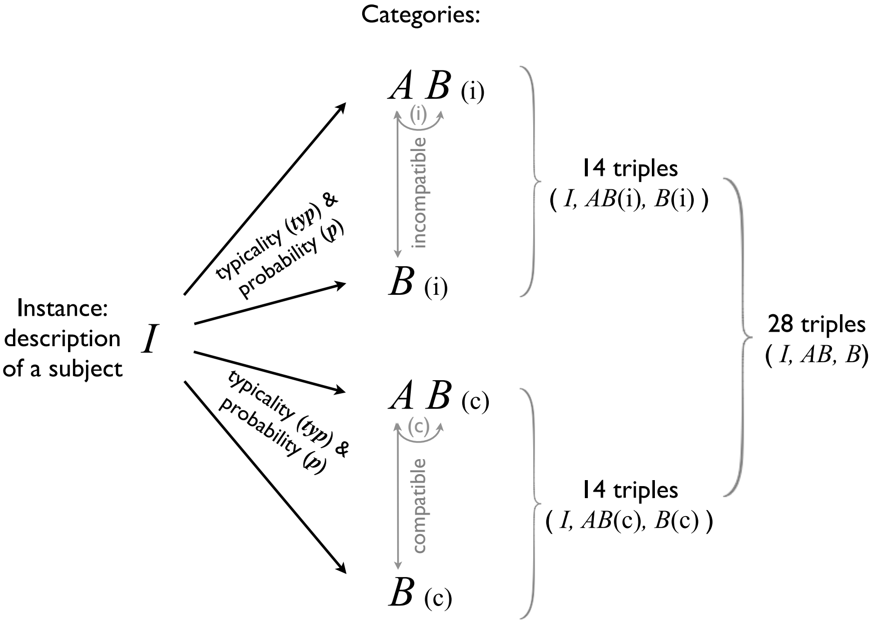

Figure 1: Experimental setup from (Shafir et al., 1990). Participants were asked to make typicality and probability judgments about 14 instances I (description of a subject) with respect to four categories: (1) a compatible conjunction ($AB(c)$), (2) its constituent $B(c)$; (3) an incompatible conjunction ($AB(i)$), and (4) its constituent $B(i)$. The compatibility (c) and incompatibility (i) characteristic were attributed by experimenters based on a qualitative assessment of a number of shared properties between conjunction elements (A and B). Judgments were made by two distinct groups: 54 decision makers for typicality, and 56 for probability.

and conjunction fallacy (with positive differences larger then 1%) were reported for 10 out of 14 instances, when coupled with *compatible* categories. The conjunction effect and conjunction fallacy were observed to be stronger and confirmed for all 14 instances, when combined with *incompatible* categories.

Table 1: Aggregate statistics of the experimental results from (Shafir et al., 1990), with a total of 28 triples that combines 14 instances I with categories AB and B, either incompatible (i) or compatible (c). Sample means μ and sample standard deviations σ for the sizes of the conjunction effect and conjunction fallacy, as well as the correlation between them, are provided. Reproduced from (Shafir et al., 1990).

	Conjunction effect: $typ_{exp}(AB) - typ_{exp}(B)$		Conjunction fallacy: $p_{exp}(AB) - p_{exp}(B)$		Correlation between conjunction effects and fallacies
	μ	σ	μ	σ	
28 triples (I,AB,B), incl.:	0.133	0.115	0.078	0.092	0.83 (p<0.01, N=28)
14 triples (I,$AB(i)$,$B(i)$)	0.204	0.063	0.126	0.064	0.58 (p<0.05, N=14)
14 triples (I,$AB(c)$,$B(c)$)	0.063	0.113	0.030	0.093	0.81 (p<0.05, N=14)

Significant positive correlation between the magnitudes of the conjunction effects and fallacies was reported. Note that the correlation is lower for the pairs involving incompatible categories.

For the full description of the experiment, corresponding instances and categories, we refer to the original study (Shafir et al., 1990).

3 Previous interpretation of the conjunction fallacy within Quantum decision theory

3.1 Summary of the previous interpretation

The previous interpretation of the conjunction fallacy within Quantum decision theory (QDT), based on the experimental results of (Shafir et al., 1990), was proposed in (Yukalov and Sornette, 2009). The current section condenses the arguments of that interpretation, and provides citations from the original.[1] Parts of sentences in italic correspond to pieces of text extracted from (Yukalov and Sornette, 2009). Within these quotes, we indicate within parentheses when we had to adapt the text to make it understandable with our present conventions.

Consider the following two intentions. *One intention, with just one representation, is "to decide whether the (su)bject* (which is described in an instance I) *has the feature B." The second intention is "to decide about the secondary feature" which has two representations, when one decides whether "the (su)bject has the special characteristic" (A_1) or "the (su)bject does not have this characteristic" (A_2).* Thus, according to the notation (B = "bankteller", A_1 = "feminist", A_2 = "non-feminist"), prospects are formulated as follows:

- for conjunction $AB(i)$ ("feminist bankteller"):

$$\pi_1 = BA_1, i.e. |\pi_1\rangle = a_{11}|BA_1\rangle, \tag{1}$$

- for its constituent $B(i)$ ("bankteller"):

$$\pi_2 = BA = B(A_1 + A_2), i.e. |\pi_2\rangle = a_{21}|BA_1\rangle + a_{22}|BA_2\rangle. \tag{2}$$

The following general scheme is applied to calculate the probability of the prospect with a constituent:

$$\begin{aligned} p(\pi_2) &= p(BA) = p(BA_1) + p(BA_2) + q(BA) = \\ &= p(B|A_1)p(A_1) + p(B|A_2)p(A_2) + q(BA) \end{aligned} \tag{3}$$

This is a typical situation where a decision is taken under uncertainty. The uncertainty-aversion principle requires that the interference term $q(BA)$ should be negative ($q(BA) < 0$).

For the set of compatible pairs of characteristics, it turned out that the average probabilities were $p(BA)$ = 0.537 and $p(BA_1)$ = 0.567, with statistical errors of 20%. Hence, within this accuracy, $p(BA)$ and $p(BA_1)$ coincide and no conjunction fallacy arises for compatible characteristics. From the view point of QDT, this is easily interpreted as due to the lack of uncertainty: since the features

---

[1] To unify notation, B is used for a characteristic that appears in both (i) a conjunction and (ii) a constituent categories (e.g. "bankteller"), and A - for a characteristic that occurs only in a conjunction (e.g. "feminist"). This notation corresponds to (Shafir et al., 1990) and replaces the corresponding symbols that were used in (Yukalov and Sornette, 2009), where A = "bankteller", X_1 = "feminist" and X_2 = "non-feminist".

B and A_1 are similar to each other, one almost certainly yielding the other, there is no uncertainty in deciding, hence, no interference, and, consequently, no conjunction fallacy.

For *incompatible* categories, *the simplest and most natural mathematical embodiment of the property of "incompatibility" is to take the probabilities of possessing B, under the condition of either having or not having A_1, as equal, that is, $p(B|A_j) = 0.5$.* For these incompatible pairs of categories, Equation (3) reduces to

$$p(BA) = \frac{1}{2} + q(BA). \tag{4}$$

For *incompatible* categories, the average values of the reported probabilities are $p(BA) = 0.220$ and $p(BA_1) = 0.346$ (Shafir et al., 1990).

Given the observed values of $p(BA)$ for each of the 14 constituents of incompatible categories (i) (Shafir et al., 1990) and Equation (4), the observed interference terms are found fluctuating around a mean of −0.28, with a standard deviation of ±0.06:

$$q(BA) = -0.28 \pm 0.06. \tag{5}$$

The conjunction error is found to be:

$$\varepsilon(BA_1) \equiv p(BA_1) - p(BA) = 0.126. \tag{6}$$

From Equation (3), the average value of $p(BA_2)$ is equal to 0.154. In addition, the proposed assumption that $p(B|A_j) = 0.5$ leads to $p(A_1) = p(BA_1)/0.5 = 0.692$, and similarly $p(A_2) = p(BA_2)/0.5 = 0.308$.

Yukalov and Sornette (2009) conclude: *QDT interprets the conjunction effect as due to the uncertainty underlying the decision, which leads to the appearance of the intention interferences. The interference of intentions is caused by the hesitation whether, under the given primary feature (B), the (su)bject possesses the secondary feature (A_1) or does not have it (A_2). The term $q(BA)$ is negative, reflecting the effect of deciding under uncertainty (according to the uncertainty-aversion principle). Quantitatively, we observe that the amplitude $|q(BA)|$ is in agreement with the QDT interference-quarter law.*

3.2 Weaknesses of the previous interpretation

As summarised in the previous section 3.1, the interpretation of the conjunction fallacy in (Yukalov and Sornette, 2009) rests on two assumptions:

1. the formulation of a prospect for a constituent category B such that it includes uncertainty about a secondary feature ($A_1 + A_2$);
2. the independence of incompatible features, which underlies equation (4).

The current section analyses these two assumptions and demonstrates the existence of some inconsistencies.

3.2.1 Formulation of the prospect for a constituent category (B): intention concerning a primary feature (equation 1) and undetermined sign of q

For a secondary characteristic (e.g. "feminist"), the uncertainty about its presence (i.e. an undecided attribution of this feature to the subject from an instance I) is represented as a *composite* action A, which is a sum of two action modes A_1 ("the subject *has* a secondary feature") and A_2 ("the subject *does not have* a secondary feature"). However, for a primary characteristic (e.g. "bankteller"), a *simple* action with one action mode B ("*to decide* whether the subject has a primary feature") is suggested. This formulation of intention B as an *active decision* concerning a primary feature is necessary to justify the negative sign of the attraction factor q for a prospect with a constituent category (B), if a level of uncertainty about a secondary feature ($A_1 + A_2$) is introduced in this prospect, such that:

$$\pi_2 = B(A_1 + A_2)\ , \text{ i.e. } \quad |\pi_2\rangle = a_{21}\,|B \otimes A_1\rangle + a_{22}\,|B \otimes A_2\rangle\ , \tag{7}$$

where the symbol $\otimes$ represents the tensor product operator (see (Yukalov and Sornette, 2009) for details).

Thus, under an assumption of passivity in the presence of uncertainty, it is proposed that making a *decision* concerning a primary feature B (e.g. "bankteller") is not attractive, i.e. $q(\pi_2) \leq 0$.

The following two arguments reveal an inconsistency in the above formulation of a constituent intention B.

First, it is natural to assume that, for a primary characteristic, an intention complementary to B should exist, which is denoted for simplicity $notB$ and stands for "*not to decide* whether the subject has a primary feature"). Action $notB$ reflects an undecided attribution of a primary feature and, similar to action A for a secondary feature, can be presented as a sum of two action modes: B_1 ("the subject has a primary feature") and B_2 ("the subject does not have a primary feature"). Continuing this analogy, in formulating the prospects for both categories, i.e. for a conjunction and its constituent, action mode B_1 should be used to represent a consideration (attribution) of a primary feature (e.g. "bankteller"). However, the introduction in (Yukalov and Sornette, 2009) of action B as described above remains unclear. In fact, an interpretation of action B as an *active decision* ("*to decide* whether the object has a primary feature") seems unrealistic, as in the experiment participants were exposed to the predefined categories and *had to* judge the corresponding probabilities, i.e. participants were not asked to decide whether to make the judgement or not.

Second, if the intention B represents a single action mode of possessing a primary feature that is equivalent to B_1 ("the subject *has* a primary feature"), then a complementary action mode B_2 ("the subject *does not have* a primary characteristic) exists and also requires an *active decision* of a decision maker about the possession (or absence) of a primary feature in the subject. Thus, the sign of the attraction factor q for a constituent category cannot be determined, even when assuming the presence of uncertainty concerning a secondary feature.

These two inconsistencies can thus be summarised as follows:

- an intention concerning a primary feature B (e.g. "bankteller") should be formulated similarly to an intention about a secondary feature A (e.g. "feminist"), in the form of "the subject *has* a feature";
- the sign of the attraction factor q for a constituent category cannot be determined, even when assuming the presence of uncertainty with respect to a secondary feature in this prospect.

3.2.2 Formulation of the prospect for a constituent category (B): uncertainty about a secondary feature (equation 2)

In order to fully understand the assumptions underlying the QDT interpretation of Yukalov and Sornette (2009), it is important to note that the experiments that have investigated the conjunction fallacy have been performed under several distinct treatments, which introduce subtle but important differences for their theoretical interpretation. The following three main classes of experiment treatment have been used.

1. *Indirect* tests, when subjects were exposed to a description of a subject (an instance I) and only one of the categories (either conjunction AB, or its constituent B) at a time. In this setup, the judgements about the probability of an instance I with respect to each category - AB or B - were made *separately* and were not juxtaposed. For example, in (Tversky and Kahneman, 1983), an indirect between-subjects comparison was conducted, when the probability of the conjunction was evaluated by one group and the probability of its constituent was evaluated by another group. In (Shafir et al., 1990), judgments of probabilities for each category - AB or B - were performed separately, but by the same decision makers.
2. *Direct-subtle* tests, when, following an instance I, participants are exposed to both a conjunction and its constituent category, but the inclusion relation is not made apparent. For example, in (Tversky and Kahneman, 1983), the two categories of interest are shown simultaneously, but are camouflaged among five additional filler items.
3. *Direct-transparent* tests, when an instance I, a conjunction and its constituent are presented together to highlight the connection between the categories.

In the current QDT treatment (Yukalov and Sornette, 2009), the conjunction fallacy, i.e. $p(AB) > p(B)$, is explained by the negative attraction factor of a prospect with constituent category B. For a negative q to appear, a judgement about the prospect with a constituent B is *assumed* to be influenced by the existence of a level of *uncertainty* about the presence of a secondary feature A. In other words, a judgment about a primary feature B is saddled with an added degree of uncertainty about the attribution of a secondary feature A, even when absent in the judged category B (equation 2).

However, in the *indirect* test design as presented in (Shafir et al., 1990) and analyzed in (Yukalov and Sornette, 2009), the judgement about the probability of an instance I with respect to a constituent category B was made *separately*, without exposition to a conjunction category AB. In this indirect setup, it is not obvious that a secondary feature, which is present only in a conjunction, has any influence on a judgment about a constituent category. Thus, there is no evidence that a secondary feature should be included in the formulation of the QDT prospect for a constituent category B (e.g. “bankteller”), which instead can be simply represented by

$$\pi_2 = BT \ \text{, i.e.} \quad |\pi_2\rangle = a_{21}\,|BT\rangle \ , \tag{8}$$

instead of equation 2.

Importantly, for *indirect* tests, with this formulation of a prospect for a constituent B, there is no uncertainty about the presence of a secondary feature A in a constituent category. Thus, the QDT uncertainty aversion principle ought not to be invoked to explain the observed conjunction fallacy, and another mechanism is required.

For the *direct* test designs, which allow for a direct comparison of a conjunction and a constituent categories, the formulation of prospects proposed in (Yukalov and Sornette, 2009) is more plausible. It could be expected though that a more profound manifestation of uncertainty for a constituent category, which is associated with the presence of a secondary feature, would increase the absolute value of a negative attraction factor of this prospect ($|\pi_2\rangle$), amplifying the conjunction fallacy. However, the opposite results are observed in experiments (Tversky and Kahneman, 1983). Probably, the most convincing evidence was obtained in a *direct-transparent* test, where the probability judgment about an instance I (the description of Linda) was made with respect to the following two categories:

1. Linda is a bank teller *whether or not* she is active in the feminist movement: $B(i)$ (versus $B(A_1 + A_2)$ in (Yukalov and Sornette, 2009));
2. Linda is a bank teller and is active in the feminist movement: $AB(i)$ (versus BA_1 (Yukalov and Sornette, 2009)).

In this example, the degree of uncertainty about the presence of a secondary "feminist" feature ($A_1 + A_2$) is made explicit in a constituent category $B(i)$, which provides a good match for the formulation of a prospect with a constituent in (Yukalov and Sornette, 2009). However, contrary to what one could expect from the formulation of Yukalov and Sornette (2009), the portion of decision makers who committed the conjunction fallacy dropped from above 80% (observed in both indirect and direct tests) to 57% (for the direct-transparent test) (Tversky and Kahneman, 1983). This finding signals that the recognition by decision makers of a level of uncertainty about a secondary feature in a constituent category does not make this prospect less attractive, but rather emphasizes the inclusive relation between the two categories (conjunction and its constituent) and facilitates the correct application of the conjunction rule.

Furthermore, Tversky and Kahneman (1983) outlined that the *representativeness heuristic* may be at the heart of the persistent conjunction fallacy. Even when provided with a valid and clear explanation of the inclusion of a conjunction category into a constituent, the majority of subjects choose to stick to an "emotional" resemblance argument.

This suggests an alternative QDT mechanism for the explanation of the conjunction fallacy: rather than a negative attraction of a constituent category due to the uncertainty of a secondary feature, the key ingredient is a *higher attraction to a conjunction prospect*, if a secondary feature is compatible with the description of an instance I.

3.2.3 Independence of (incompatible-type) prospects (equation 4)

Equation (4), which is a key result in (Yukalov and Sornette, 2009), is based on two underlying assumptions:

- the "incompatibility" of constituents in a conjunction category is treated as leading to their "independence", i.e. the probability of possessing a primary characteristic B is assumed to be independent from having a secondary characteristic A_j, yielding $p(B|A_1) = p(B|A_2) = p(B)$;
- with no prior information, it is assumed that $p(B) = 0.5$.

A first general criticism can be raised about the assumption that the existence of uncertainty about an independent category would lead to a negative attraction factor when deciding about another independent category. If this was the case, any decision would then be associated with $q < 0$, as it is impossible to create a completely certain environment in our complex uncertain world, as there are always many variables that remain uncertain around us.

Secondly, in (Yukalov and Sornette, 2009), the "compatibility" of categories is associated with an absence of uncertainty, i.e. the possession of one of the compatible characteristics yields the other one, which implies positive correlation close to 1. In this context, “incompatibility” is more likely to imply negative correlation close to −1, rather than independence with correlation close to 0.

Thirdly, an assumption that a high degree of “compatibility” (or “incompatibility”) is associated with very low uncertainty, and thus leads to $q \to 0$, has to be tested. For example, it is plausible that the subjective estimation of a high “compatibility” (“incompatibility”) of categories may increase the ‘subjective’ confidence of a decision maker, which could be reflected in a high positive (negative) value of the attraction factor, and make a choice deviate from an ‘objective’ judgement. Empirical evidence should be gathered to support this hypothesis.

We are thus led to suggest two alternative propositions replacing the two assumptions of Yukalov and Sornette (2009) discussed above:

- Independent intentions do not interfere. Thus, the existence of uncertainty about one intention does not influence the probability that another intention will be realized, if these intentions are independent.
- Equation (4) requires revision: if (incompatible) intentions are treated as independent, they should not interfere and $q = 0$; if (incompatible) intentions interfere and $q \neq 0$, then $p(B|A_1) = p(B|A_2) = p(B) = 0.5$ can not be assumed.

3.2.4 Partial use of data

Yukalov and Sornette (2009) did not make use of experimental data on *compatible* conjunctions and their constituents, and on *typicality* judgements. Most importantly, the description of a subject - instance I - is a key element of the experiment, which consists in framing participants prior the choice (judgement). However, this was ignored in many theoretical interpretations, including (Yukalov and Sornette, 2009).

3.3 Synthesis

The previous interpretation of the conjunction fallacy within QDT (Yukalov and Sornette, 2009) aimed at explaining it for the most clearcut cases of one type of incompatible prospects with a constituent category $B(i)$. Agreement between partial empirical data and the general QDT relation $p{=}f{+}q$ was obtained. However, the needed underlaying assumptions have been shown to be unsustantiated.

In particular, the representation of prospects, which is needed to justify the application of the uncertainty-aversion principle and the corresponding negative sign of the attraction factor q, leads to serious inconsistencies. The interference effects argued to occur for a single constituent category B, as formulated in (Yukalov and Sornette, 2009), have a shaky foundation. As discussed above in details, within an indirect test setup (Shafir et al., 1990), the inclusion of uncertainty about a secondary feature A from a conjunction category AB should not be relevant to a separate judgement regarding B. Another essential assumption about the independence of incompatible categories, upon which the proposed interpretation rests, is arbitrary. Importantly, the influence of framing, i.e. the pre-exposure of decision makers to an instance I (the description of a subject), is disregarded. Last but not least, the available empirical data is used only partially (just for one out of four judged categories).

Since the attempt of Yukalov and Sornette (2009), the theoretical construct of QDT has been significantly enriched. We use this opportunity to ‘cure’ the above mentioned weaknesses, to propose a genuine rationalisation and quantitative explanation of the conjunction fallacy within QDT, which allows us to further explore the limits of QDT.

4 Novel theoretical reinterpretation of the conjunction fallacy with QDT

The mean idea to explain the conjunction fallacy is that decision modes of prospects and of a decision-maker's state of mind are mutually related. The probability judgement, i.e. prospect probability, is influenced by the interfering decision modes of the state of mind, which were intentionally incepted by a specific description of a subject (e.g., an instance I as a framing tool). The description of a subject is represented as an uncertainty union, due to the incompleteness of the description (a subject is briefly characterized by just a few features) and the large uncertainty in the evolution of the mentioned characteristics. In typical experiments investigating the conjunction fallacy, the decision maker is not confronted with a task of specifying a definitive set of a subjects' characteristics. These attributes remain undefined and are represented by an intermediate inconclusive event with interfering modes.

In the following subsections, we first introduce QDT concepts to characterize a decision-maker and possible events (decisions). Then we describe a decision making process with a two-step methodology and demonstrate the origin of interference effect that can yield the conjunction fallacy.

4.1 The strategic decision-maker state

In QDT, a decision maker is characterized by a strategic decision-maker state, which is in general represented by a statistical operator $\hat{\rho}$ on a decision-maker space of mind $\mathcal{H}$. A space of mind is a Hilbert space that is spanned by a set of orthogonal basic states $\{|e\rangle\}$ - all admissible events or decisions, which are considered by a decision maker. A strategic decision-maker state $\hat{\rho}$ defines the probabilities of prospective decisions to be taken. It reflects individuality (e.g. persistent personality traits) interconnected with a surrounding (e.g. memory, experience, social influence, as well as fleeting impressions) in the context of a specific choice situation.

A strategic decision-maker state evolves over time. This means that the representation basis as well as coefficients of decomposition are time-dependent. Basic personality traits and fundamental values are relatively stable and undergo a gradual transformation. Weights of these inherent individual characteristics are predominant for choices involving important outcomes, such as life-determining events that require thorough deliberation. However, most of everyday choices, as well as experimental setups, are concerned with relatively minor outcomes and are subjected to a time constraint. Thus, in many situations, simplified rules, i.e. heuristics, are employed to make decisions. Depending on the task at hand, only a few factors come to the fore (e.g. more recent, frequent or typical features) and substantially determine the choice. Practically, this suggests that, in certain cases, a strategic decision-maker state $\hat{\rho}$ can be reduced to a few contextually dependent dimensions (decision modes), or (and) even represented by a pure state $\hat{\rho} = |\psi\rangle\langle\psi|$.

Thus, the description of a choice situation (e.g. a general experimental setup, the formulation of options) are decisive for the basis composition and the resultant prospect probability. Measurable deviations in choice can be generated by intentional manipulations with a task context and produce specific behavioral patterns, such as framing or anchoring effects.

4.2 Decision modes of the experiment

At the beginning of the experiment, before any category for a probability judgment is formulated, a decision maker is pre-exposed to an instance that describes a subject by a combination of characteristics. Let us introduce an observable I with a set of possible values:

$$I = \{I_1, I_2\} = \biguplus_k I_k \qquad (k = 1, 2), \tag{9}$$

where I_1 reflects characteristics that are attributed to a subject from a description, while I_2 includes characteristics that do not fit the image presented in an instance, i.e. I_1 and I_2 represent complementary characteristics. Thus, specific features are highlighted in the mind of a decision maker, and contrasted to others. This classification facilitates certain operations, such as recognition of a subject and its comparison with other subjects, by matching their characteristics. However, a high degree of uncertainty in the attribution of features to a subject remains due to the brevity and fuzziness of a description. Firstly, the classification can be ambiguous for a characteristic that is resembling other qualities of a subject and is typical of him/her, but is not directly mentioned in the presented description. Secondly, the strength and evolution of individual inclinations can be obscure. For example, early interests can be developed to a professional level or dissipated over time. Until the related uncertainty is not resolved in the mind of a decision maker by an observable decision, i.e. a specific value I_k has not been chosen, I is considered to be an inconclusive (or operationally uncertain) event, and the set (9) is an uncertain union (Yukalov and Sornette, 2015).

The state vector $|I\rangle$ corresponding to the observable I is thus a linear combination

$$|I\rangle = \gamma_1 |I_1\rangle + \gamma_2 |I_2\rangle \tag{10}$$

of mode states $|I_k\rangle$ (k=1,2) with probability weights $|\gamma_k|^2$. The set of vectors $\{|I_1\rangle, |I_2\rangle\}$ forms an orthonormal basis

$$\langle I_k | I_l \rangle = \delta_{kl}, \quad \text{where } \delta_{kl} \text{ is the Kronecker delta,} \tag{11}$$

and its linear span is a Hilbert space $\mathcal{H}_I$

$$\mathcal{H}_I \equiv span\{|I_1>, |I_2>\}. \tag{12}$$

For an uncertain union (9), a state vector (10) generates an operator $\hat{P}_I$, which forms an operator algebra $\mathcal{I} \equiv \{\hat{P}_{I_k}\}$ and is expressed as

$$\begin{aligned} \biguplus_k I_k \quad \to \quad \hat{P}_I \quad \equiv \quad |I\rangle\langle I| \quad &= \quad \sum_k |\gamma_k|^2 |I_k\rangle\langle I_k| + \sum_{k \neq l} \gamma_k \gamma_l^* |I_k\rangle\langle I_l| \\ &= \quad \sum_k |\gamma_k|^2 \hat{P}_{I_k} + \sum_{k \neq l} \gamma_k \gamma_l^* |I_k\rangle\langle I_l| \qquad k, l \in \{1, 2\}. \end{aligned} \tag{13}$$

In the last summand of (13), the term $|I_k\rangle\langle I_l|$, for $k \neq l$, does not belong to the operator algebra $\mathcal{I}$. In other words, the operator $\hat{P}_I$ cannot be represented as a linear combination of the elements from its algebra $\{\hat{P}_{I_k}\}$. Thus, the operator $\hat{P}_I$ is called entangled, and the modes I_k of the corresponding inconclusive event I are interfering with each other (Yukalov and Sornette, 2015).

In indirect tests, the description of a subject (an instance I) is followed by one of the categories, either a conjunction $A\&B$, or its constituent B. Let us introduce two observables AB and B with the sets of their possible values:

$$AB = \{(AB)_1, (AB)_2\} = \bigcup_i (AB)_i \qquad (i = 1, 2) \quad \text{and} \tag{14}$$

$$B = \{B_1, B_2\} = \bigcup_i B_i \qquad (i = 1, 2), \tag{15}$$

where $(AB)_1$ (resp., B_1) represents characteristics of a conjunction (resp., constituent) category, or attribution of these features to a subject when making a judgement about him/her. Then $(AB)_2$ (resp., B_2) is a complementary subevent (mode), when a conjunction (resp., constituent) category is considered to be absent, or relevant features are not attributed to a subject.

In the course of the experiment, a participant makes an explicit probability judgment about attribution of a conjunction category or its constituent to a described subject. Thus, both events - AB and B - are operationally testable, i.e. each of the observables takes a concrete value from a corresponding set, either (14) or (15). This consideration underpins two assumptions: (i) sets of observable values are represented by standard unions; (ii) a conjunction is treated as a single category AB, i.e. a tensor product of two constituent categories, such that $AB = A \otimes B$.

For both observables AB and B, we put into correspondence the state vectors $|AB\rangle$ and $|B\rangle$ that are decomposed onto orthonormal bases as follows:

$$|AB\rangle = \alpha_1 \,|(AB)_1\rangle + \; \alpha_2 \,|(AB)_2\rangle \,, \quad \langle (AB)_i | (AB)_j \rangle = \delta_{ij}, \tag{16}$$

$$|B\rangle = \beta_1 \,|B_1\rangle + \; \beta_2 \,|B_2\rangle \,, \quad \langle B_i | B_j \rangle = \delta_{ij}, \tag{17}$$

with respective probability weights $|\alpha_i|^2$ and $|\beta_i|^2$ (i=1,2). The vectors of the bases - $\{|(AB)_i\rangle\}$ and $\{|B_i\rangle\}$ - are spanning sets of the corresponding Hilbert spaces $\mathcal{H}_{AB}$ and $\mathcal{H}_B$:

$$\mathcal{H}_{AB} \equiv span\{|(AB)_1\rangle \,, |(AB)_2\rangle\}, \tag{18}$$

$$\mathcal{H}_{B} \equiv span\{|B_1\rangle \,, |B_2\rangle\}. \tag{19}$$

The operators $\hat{P}_{AB}$ and $\hat{P}_B$ are generated by the respective state vectors (16) and (17):

$$\bigcup_i (AB)_i \quad \rightarrow \quad \hat{P}_{AB} \;\equiv\; \sum_i \hat{P}_{(AB)_i} \;=\; \sum_i |(AB)_i\rangle \langle (AB)_i| \qquad i = 1, 2. \tag{20}$$

$$\bigcup_i B_i \quad \rightarrow \quad \hat{P}_{B} \;\equiv\; \sum_i \hat{P}_{B_i} \;=\; \sum_i |B_i\rangle \langle B_i| \qquad i = 1, 2. \tag{21}$$

Note that, because the operators $\hat{P}_{AB}$ and $\hat{P}_B$ correspond to operationally testable events, they can be decomposed into elements of their operator algebras $\mathcal{AB} = \{\hat{P}_{(AB)_i}\}$ and $\mathcal{B} = \{\hat{P}_{B_i}\}$, as in (20) and (21), i.e. they are not entangled. Thus, decision modes within each of the corresponding operationally testable events AB and B are not interfering (Yukalov and Sornette, 2015).

In a typical quantum measurement, an experiment with a physical system consists of two phases: preparation of a system state and measurement of an observable. Decision making can also be described as a two-step process: first, *deliberation* about objectives, desires, choice alternatives, constraints, and so on, and, second, *taking a decision* by adopting a certain choice option. Within QDT, this translates into, (i) an initial preparation and evolution of a strategic decision-maker state, influenced by the context of the choice situation, and (ii) a convergence to and observation of a concrete event (a decision) out of a set of possible basic decision modes. A probabilistic interpretation implies that an event (a decision) occurs with a certain probability that can be predicted from the state of mind of the decision maker.

In the following subsections, these two phases of a decision making process are described with two types of representation: subsection 4.3 uses a composite state representation, which is convenient for representing the measurement, i.e., the observation of a decision; subsection 4.4 applies a channel representation, which is more suitable to follow the evolution of a strategic decision-maker state during an initial preparation phase.

4.3 General representation of a strategic decision-maker state of mind with statistical operators

4.3.1 Preparation: evolution of a space of mind and of a strategic decision-maker state

In the first step of a preparation phase, a decision maker is exposed to the description of a subject (an instance I). An initial space of mind $\mathcal{H}_M$ is enlarged to a Hilbert-space tensor product $\mathcal{H}_{IM}$

$$\mathcal{H}_M \quad \to \quad \mathcal{H}_{IM} \equiv \mathcal{H}_I \otimes \mathcal{H}_M. \tag{22}$$

The evolution over the time interval $[t_0; t_1]$ of a strategic decision-maker state $\hat{\rho}_M(t_0)$, influenced by an introduced partial state $\hat{\rho}_I(t_0)$, is represented with the entangling channel

$$C_1: \quad \hat{\rho}_I(t_0) \otimes \hat{\rho}_M(t_0) \quad \to \quad \hat{\rho}_{IM}(t_1). \tag{23}$$

Between intentional external inputs, the space of mind $\mathcal{H}_{IM}$ is assumed to be unchanged. However, the strategic state of a decision-maker can evolve due to internal processes of deliberation as well as the influence of the environment in the time interval $[t_1; t_2]$. This is captured by a unitary evolution operator $\hat{U}$:

$$C_2: \quad \hat{\rho}_{IM}(t_1) \quad \to \quad \hat{\rho}_{IM}(t_2) = \hat{U}(t_2 - t_1)\hat{\rho}_{IM}(t_1)\hat{U}^\dagger(t_2 - t_1) \tag{24}$$

The introduction of a new category over the time interval $[t_2; t_3]$ - a conjunction $A\&B$ (resp., its constituent B) - expands the space of mind

$$\begin{aligned} \mathcal{H}_{IM} \quad &\to \quad \mathcal{H}_{ABIM} \equiv \mathcal{H}_{AB} \otimes \mathcal{H}_{IM} \\ (resp., \quad \mathcal{H}_{IM} \quad &\to \quad \mathcal{H}_{BIM} \equiv \mathcal{H}_B \otimes \mathcal{H}_{IM}), \end{aligned} \tag{25}$$

and a decision-maker state is further entangled through the channel

$$\begin{aligned} C_3: \quad \hat{\rho}_{AB}(t_2) \otimes \hat{\rho}_{IM}(t_2) \quad &\to \quad \hat{\rho}_{ABIM}(t_3) \\ (resp., \quad C3: \quad \hat{\rho}_B(t_2) \otimes \hat{\rho}_{IM}(t_2) \quad &\to \quad \hat{\rho}_{BIM}(t_3)). \end{aligned} \tag{26}$$

An intermediate evolution over the time interval $[t_3; t_4]$ with a unitary operator $\hat{U}$ completes the preparation of a strategic decision-maker state, which is entangled with a conjunction category AB (resp., its constituent B):

$$\begin{aligned} C_4: \quad \hat{\rho}_{ABIM}(t_3) \quad &\to \quad \hat{\rho}_{ABIM}(t_4) = \hat{U}(t_4 - t_3)\hat{\rho}_{ABIM}(t_3)\hat{U}^\dagger(t_4 - t_3) \\ (resp., \quad C_4: \quad \hat{\rho}_{BIM}(t_3) \quad &\to \quad \hat{\rho}_{BIM}(t_4) = \hat{U}(t_4 - t_3)\hat{\rho}_{BIM}(t_3)\hat{U}^\dagger(t_4 - t_3). \end{aligned} \tag{27}$$

The attribution to a subject (described in an instance I) of a conjunction category $A\&B$ (resp., its constituent B) is an operationally testable event, which is associated with an explicit decision (similar to a measurement of a quantum system) and can be observed with a certain probability. In general, the procedure of making a decision (measurement) can be described by partially disentangling channels that lead to a separation of $\hat{\rho}_I$ and $\hat{\rho}_{AB}$ (resp., $\hat{\rho}_B$) from the total decision-maker state $\hat{\rho}_{ABIM}$ (resp., $\hat{\rho}_{BIM}$). This description is realistic, but involves several additional entangling-disentangling channels.

However, as demonstrated in (Yukalov and Sornette, 2016a), the channel-state duality established by the Choi-Jamiolkowski isomorphism allows one to equivalently represent a measurement by the introduction of a composite state. Thus, for the second phase of the decision making processes (revelation of a decision), a channel representation is conveniently substituted by a composite state representation.

4.3.2 Measurement: decision under uncertainty (composite prospect and its probability)

Within QDT, taking a decision (adopting an option) is equivalent to a transition of the strategic state of a decision-maker to a state corresponding to the chosen option.

In the analyzed experimental setup, a probability judgment is made with respect to the attribution of a conjunction category AB (resp., its constituent B) to a subject (from an instance I). The corresponding choice option is a composite prospect of the form $AB \otimes I$ (resp., $B \otimes I$). Taking into account that a judgment about the absence of a feature, i.e. $(AB)_2$ (resp., B_2), is not investigated, a choice prospect includes only a decision mode $(AB)_1$ (resp., B_1) and is formulated as follows:

$$\begin{aligned} \pi_{(AB)_1 I} &= (AB)_1 \otimes \biguplus_{k=1,2} I_k, \\ (resp., \quad \pi_{B_1 I} &= B_1 \otimes \biguplus_{k=1,2} I_k \ ), \end{aligned} \tag{28}$$

with a corresponding prospect state $\left|\pi_{(AB)_1 I}\right\rangle$ (resp., $|\pi_{B_1 I}\rangle$):

$$\begin{aligned} \left|\pi_{(AB)_1 I}\right\rangle &= |(AB)_1\rangle \otimes (\gamma_1 |I_1\rangle + \ \gamma_2 |I_2\rangle) = \sum_{k=1,2} \gamma_k |(AB)_1 I_k\rangle \\ (resp., \quad |\pi_{B_1 I}\rangle &= |B_1\rangle \otimes (\gamma_1 |I_1\rangle + \ \gamma_2 |I_2\rangle) = \sum_{k=1,2} \gamma_k |B_1 I_k\rangle \ \ ). \end{aligned} \tag{29}$$

The basic decision modes of the prospect state $\{|(AB)_1 I_k\rangle\}$ (resp.,$\{|B_1 I_k\rangle\}$) are vectors of an orthogonal basis of the space of mind $\mathcal{H}_{ABIM}$ (resp., $\mathcal{H}_{BIM}$) given by (25).

A composite prospect state $\pi_{(AB)_1 I}$ (resp., $\pi_{B_1 I}$) generates a prospect operator

$$\begin{aligned} \pi_{(AB)_1 I} \quad \rightarrow \quad \hat{P}_{(AB)_1 I} &\equiv \left|\pi_{(AB)_1 I}\right\rangle \left\langle \pi_{(AB)_1 I}\right| \\ &= \sum_k |\gamma_k|^2 \, |(AB)_1 I_k\rangle \langle (AB)_1 I_k| + \sum_{k \neq l} \gamma_k \gamma_l^* \, |(AB)_1 I_k\rangle \langle (AB)_1 I_l| \\ &= \sum_k |\gamma_k|^2 \hat{P}_{(AB)_1} \otimes \hat{P}_{I_k} + \sum_{k \neq l} \gamma_k \gamma_l^* \hat{P}_{(AB)_1} \otimes |I_k\rangle \langle I_l| \qquad k, l \in \{1, 2\} \end{aligned} \tag{30}$$

$$\begin{aligned} (resp., \quad \pi_{B_1 I} \quad \rightarrow \quad \hat{P}_{B_1 I} &\equiv |\pi_{B_1 I}\rangle \langle \pi_{B_1 I}| \\ &= \sum_k |\gamma_k|^2 \, |B_1 I_k\rangle \langle B_1 I_k| + \sum_{k \neq l} \gamma_k \gamma_l^* \, |B_1 I_k\rangle \langle B_1 I_l| \\ &= \sum_k |\gamma_k|^2 \hat{P}_{B_1} \otimes \hat{P}_{I_k} + \sum_{k \neq l} \gamma_k \gamma_l^* \hat{P}_{B_1} \otimes |I_k\rangle \langle I_l| \qquad k, l \in \{1, 2\} \ ). \end{aligned} \tag{31}$$

The probability that a decision maker in a strategic state of mind $\hat{\rho}_{ABIM}$ (resp., $\hat{\rho}_{BIM}$) will choose the prospect $\pi_{(AB)_1 I}$ (resp., $\pi_{B_1 I}$) is given by

$$\begin{aligned} p(\pi_{(AB)_1 I}) &= \mathrm{Tr}\hat{\rho}_{ABIM}\hat{P}_{(AB)_1 I} \\ (resp., \quad p(\pi_{B_1 I}) &= \mathrm{Tr}\hat{\rho}_{BIM}\hat{P}_{B_1 I}) \end{aligned} \tag{32}$$

Its explicit form reads

$$\begin{aligned} p(\pi_{(AB)_1 I}) &= \sum_{kl}\gamma_k\gamma_l^* \langle (AB)_1 I_l|\hat{\rho}_{ABIM}|(AB)_1 I_k\rangle \\ (resp., \quad p(\pi_{B_1 I}) &= \sum_{kl}\gamma_k\gamma_l^* \langle B_1 I_l|\hat{\rho}_{ABIM}|B_1 I_k\rangle), \end{aligned} \tag{33}$$

where $k, l \in \{1, 2\}$, which can be expressed as the sum of two parts

$$p(\pi_{(AB)_1 I}) = f(\pi_{(AB)_1 I}) + q(\pi_{(AB)_1 I}) \ , \tag{34}$$

where the diagonal term, i.e. utility factor, reads

$$\begin{aligned} f(\pi_{(AB)_1 I}) &= \sum_{k}|\gamma_k|^2 \langle (AB)_1 I_k|\hat{\rho}_{ABIM}|(AB)_1 I_k\rangle \\ (resp., \quad f(\pi_{B_1 I}) &= \sum_{k}|\gamma_k|^2 \langle B_1 I_k|\hat{\rho}_{ABIM}|B_1 I_k\rangle \end{aligned} \tag{35}$$

and the off-diagonal term, called "attraction factor" by Yukalov and Sornette (2009), takes the form

$$\begin{aligned} q(\pi_{(AB)_1 I}) &= \sum_{k\neq l}\gamma_k\gamma_l^* \langle (AB)_1 I_l|\hat{\rho}_{ABIM}|(AB)_1 I_k\rangle \\ (resp., \quad q(\pi_{B_1 I}) &= \sum_{k\neq l}\gamma_k\gamma_l^* \langle B_1 I_l|\hat{\rho}_{ABIM}|B_1 I_k\rangle \end{aligned} \tag{36}$$

By construction, p and f are probabilities that take values in $[0, 1]$. Yukalov and Sornette (2009) proved the "alternation law", which states that the sum of the q-factors (which reads $q(\pi_{(AB)_1 I}) + q(\pi_{(AB)_2 I})$ according to (36) in the present binary case) is identically zero. Thus, in such binary decisions involving the presence or absence of a trait, it is sufficient to discuss the q-factor of just one of the alternatives, the other one being by construction of equal amplitude and opposite sign. Moreover, using non-informative prior assumptions, the typical amplitude of the q-factor for binary choices can be shown to equal 0.25, which is referred to as the "quarter law" (Yukalov and Sornette, 2009) (see Yukalov and Sornette (2016b) for generalisation to multiple choices beyond binary ones). Finally, from the structure (34), the contraints $p, f \in [0, 1]$ and the alternation law, Vincent et al. (2017) showed that the q-factor obeys the additional constraint

$$|q(\pi_{(AB)_1 I})| \leq \min[f(\pi_{(AB)_1 I}), 1 - f(\pi_{(AB)_1 I})] \ . \tag{37}$$

Interferences between decision modes, which is captured by a non-zero attraction factor (36), have a profound influence on the probability of a prospect to be chosen and may even reverse a decision-maker's preference (in a sense of changin the most probable choice option). As shown in Appendix A.2, the proposed QDT interpretation of the experimental setup complies with the necessary conditions for the appearance of the attraction factor.

To explain the conjunction fallacy, i.e. $p(\pi_{(AB)_1 I}) > p(\pi_{B_1 I})$, one should analyze in-depth the values of the coefficients γ in (33), (35) and (36). For this, we use a simplified representation of the strategic decision-maker state in terms of a pure state.

4.4 Simplified representation of a strategic decision-maker state as a pure state

4.4.1 Preparation: entanglement of a pure state of mind

As already been mentioned in subsection 4.1, in certain choice situations, it can be sufficiently realistic and operationally convenient to represent a strategic decision-maker state $\hat{\rho}$ in a simplified form, i.e. as a pure state $\hat{\rho} = |\psi\rangle\langle\psi|$. For example, such situations may involve time restrictions, when a thorough deliberation is not possible, or specific setups, when limited attention resources are focused sharply on a quasi-isolated task at hand. To some degree, both of these features can be attributed to short laboratory experiments as reported in (Shafir et al., 1990), which is analyzed here. Thus, in this section a decision maker is characterized by a state vector $|\psi\rangle$, and is assumed to be isolated from external and internal influences that are not explicitly formulated in the experiment.

The decision modes of the experiment remain the same as formulated in subsection 4.2.

We assume that, in the first phase of the experiment (period $[t_0; t_1]$), a decision maker concentrates her full attention on the description of a subject (an instance I). Thus, his/her mind space $\mathcal{H}_M$ converges initially to $\mathcal{H}_I$

$$\mathcal{H}_M \quad \rightarrow \quad \mathcal{H}_I \ \equiv \ span\{|I_k\rangle\}, \ k \in 1, 2. \tag{38}$$

The corresponding focused state of mind $|\psi_I\rangle$ can be represented in the Hilbert space $\mathcal{H}_I$ as a linear combination of elementary state vectors from the basis $\{|I_k\rangle\}, k = 1, 2$:

$$|\psi_I\rangle(t_1) = \zeta_1 |I_1\rangle + \ \zeta_2 |I_2\rangle\,. \tag{39}$$

The time-dependent coefficients $|\zeta_k|^2$ in (39) play a role similar to the weights $|\gamma_k|^2$ of the mode states $|I_k\rangle$ in (10). However, the state of mind vector is assumed to be normalized, which implies an additional constraint on the coefficients $|\zeta_k|^2$:

$$\langle\psi_I|\psi_I\rangle := 1 \quad \Rightarrow \quad \sum_k |\zeta_k|^2 = 1 \quad (k = 1, 2). \tag{40}$$

Thus, in this case, $|\zeta_k|^2$ and $|\gamma_k|^2$ are proportional, but not necessarily equivalent to each other.

In the course of the experiment (period $[t_1; t_2]$), the description of a subject (an instance I) is followed by a new category - a conjunction $A\&B$ (resp., its constituent B). Due to the exposition to the category, the decision-maker's mind space $\mathcal{H}_\mathcal{I}$ expands and can be represented as the Hilbert-space tensor product

$$\begin{aligned} \mathcal{H}_I \quad &\rightarrow \quad \mathcal{H}_{ABI} \ \equiv \ \mathcal{H}_{AB} \otimes \mathcal{H}_I \ \equiv \ span\{|(AB)_i I_k\rangle\}, \ i, k \in 1, 2 \\ (resp., \quad \mathcal{H}_I \quad &\rightarrow \quad \mathcal{H}_{BI} \ \equiv \ \mathcal{H}_B \otimes \mathcal{H}_I \ \equiv \ span\{|B_i I_k\rangle\}, \ i, k \in 1, 2). \end{aligned} \tag{41}$$

The expanded state of mind $|\psi_{ABI}\rangle$ (resp., $|\psi_{BI}\rangle$) is a linear combination of elementary prospects $\{|(AB)_i I_k\rangle\}$ (resp., $\{|B_i I_k\rangle\}$):

$$\begin{aligned} |\psi_{ABI}\rangle \quad &= \quad \zeta_{11}|(AB)_1 I_1\rangle + \zeta_{12}|(AB)_1 I_2\rangle + \zeta_{21}|(AB)_2 I_1\rangle + \zeta_{22}|(AB)_2 I_2\rangle \\ (resp., \quad |\psi_{BI}\rangle \quad &= \quad \kappa_{11}|B_1 I_1\rangle + \kappa_{12}|B_1 I_2\rangle + \kappa_{21}|B_2 I_1\rangle + \kappa_{22}|B_2 I_2\rangle). \end{aligned} \tag{42}$$

The elementary prospects form an orthonormalised basis, thus squared coefficients ζ sum to 1:

$$\begin{aligned} \langle\psi_{ABI}|\psi_{ABI}\rangle \quad &:= \quad 1 \quad \Rightarrow \quad \sum_{i,k} |\zeta_{ik}|^2 = 1, \quad i, k \in \{1, 2\} \\ (resp., \quad \langle\psi_{BI}|\psi_{BI}\rangle \quad &:= \quad 1 \quad \Rightarrow \quad \sum_{i,k} |\kappa_{ik}|^2 = 1, \quad i, k \in \{1, 2\}). \end{aligned} \tag{43}$$

Taking into account the definitions of decision modes of subsection 4.2, the interpretation of the elementary prospects (decision modes) may help to reverse-engineer the probability amplitudes ζ (resp., κ), as follows.

- $|(AB)_1 I_1\rangle$ (resp., $|B_1 I_1\rangle$) implies an attribution of the features of a category AB (resp., B) to a subject that simultaneously possesses features from the description I. Thus, this vector can be associated with compatible features of both, a judged category AB (resp., B) and an instant I. If category and description of a subject share many common features (i.e. AB (resp., B) and I are *compatible*), then one could expect larger values of the probability amplitude ζ_{11} (resp., κ_{11}) of the corresponding elementary prospect, and consequently higher judged typicality/probability of AB (resp., B) with respect to I.
- $|(AB)_1 I_2\rangle$ and $|(AB)_2 I_1\rangle$ (resp., $|B_1 I_2\rangle$ and $|B_2 I_1\rangle$) reflect modes of simultaneous attribution to a subject of either (i) features associated with a judged category AB (resp., B) and features complementary to (i.e. disassociated with) a subject's initial description I; or vice versa (ii) features complementary to (i.e. disassociated with) a judged category AB (resp., B) and features that are compliant with an instance I. Thus, a stronger dissimilarity between a judged category and the description of a subject should increase the probability amplitudes ζ_{ik} (resp., κ_{ik}) $(i \neq k)$, which are associated with *incompatibility* of AB (resp., B) and I.
- $|(AB)_2 I_2\rangle$ (resp., $|B_2 I_2\rangle$) represents a state when a subject is endowed with *complementary* features of both a judged category AB (resp., B) and an instance I. This mode corresponds to the judgement of a decision maker that neither a description from an instance I, nor features associated with a category AB (resp., B), can be attributed to the subject under consideration.

Here, we propose that, when a decision maker is focused on an experimental task - i.e. is exposed to the description of a subject I, a category AB (resp., B) and attempts to judge typicality/probability of the latter - he/she concentrates almost exclusively on compatible and incompatible features of the two. Because the state of mind is focused on the presence of specific characteristics, the possibility of the considered subject to deviate from both the description of an instance I and a judged category AB (resp., B), is largely disregarded. Thus, in the notation used above, we suggest that $\zeta_{22} \to 0$ (resp., $\kappa_{22} \to 0$). This assumption evokes employing representativeness and availability heuristics, while neglecting a base rate in judgements.

Taking into account, as discussed above, that $\zeta_{22} \to 0$ (resp., $\kappa_{22} \to 0$) and other coefficients ζ_{ik} (resp, κ_{ik}) $(i, k \in 1, 2)$ are non-zero, the state of mind (42) becomes

$$\begin{aligned} |\psi_{ABI}\rangle &= \zeta_{11}\,|(AB)_1 I_1\rangle + \zeta_{12}\,|(AB)_1 I_2\rangle + \zeta_{21}\,|(AB)_2 I_1\rangle \\ (resp., \quad |\psi_{BI}\rangle &= \kappa_{11}\,|B_1 I_1\rangle + \kappa_{12}\,|B_1 I_2\rangle + \kappa_{21}\,|B_2 I_1\rangle). \end{aligned} \tag{44}$$

Note that (44) is an *entangled* state of mind, because it cannot be represented in a separable form, as the tensor product of the elementary (intention) states,

$$\begin{aligned} |\psi_{ABI}\rangle &= |AB\rangle \otimes |I\rangle = \alpha_1\gamma_1\,|(AB)_1 I_1\rangle + \alpha_1\gamma_2\,|(AB)_1 I_2\rangle + \alpha_2\gamma_1\,|(AB)_2 I_1\rangle + \alpha_2\gamma_2\,|(AB)_2 I_2\rangle \\ (resp., \quad |\psi_{BI}\rangle &= |B\rangle \otimes |I\rangle = \beta_1\gamma_1\,|B_1 I_1\rangle + \beta_1\gamma_2\,|B_1 I_2\rangle + \beta_2\gamma_1\,|B_2 I_1\rangle + \beta_2\gamma_2\,|B_2 I_2\rangle)\ , \end{aligned} \tag{45}$$

in the sense that there exists no set of parameters $\{\alpha_1, \alpha_2, \gamma_1, \gamma_2\}$ (resp., $\{\beta_1, \beta_2, \gamma_1, \gamma_2\}$) that can be matched to the set of parameters $\{\zeta_{11}, \zeta_{12}, \zeta_{21}, \zeta_{22} = 0\}$ (resp., $\{\kappa_{11}, \kappa_{12}, \kappa_{21}, \kappa_{22} = 0\}$).

Thus, prior to taking a decision, i.e. a probability judgement, a decision maker is in the entangled state of mind (44). The coefficients $|\zeta_{11}|^2$, $|\zeta_{12}|^2$, $|\zeta_{21}|^2$ (resp., $|\kappa_{11}|^2$, $|\kappa_{12}|^2$, $|\kappa_{21}|^2$) of this decision maker's state of mind are associated with a degree of compatibility of I and AB (resp., B), or with a range of shared properties between the description of a subject and a category. Herewith, $|\zeta_{11}|^2$ (resp., $|\kappa_{11}|^2$) reflects the commonality of the two, while $|\zeta_{12}|^2$ and $|\zeta_{21}|^2$ (resp., $|\kappa_{12}|^2$ and $|\kappa_{21}|^2$) quantify their distinctive unshared properties. Thus, we propose that typicality judgements, which were revealed during the experiment, can be used to quantify the coefficients of the state of mind. Taking into account equation (43), this leads us to parameterise them under the form

$$\begin{cases} |\zeta_{11}|^2 = typ(AB) \\ |\zeta_{12}|^2 = \omega(1 - typ(AB)) \\ |\zeta_{21}|^2 = (1-\omega)(1 - typ(AB)) \end{cases} \left(\text{resp.,} \begin{cases} |\kappa_{11}|^2 = typ(B) \\ |\kappa_{12}|^2 = \omega(1 - typ(B)) \\ |\kappa_{21}|^2 = (1-\omega)(1 - typ(B)) \end{cases} \right) \tag{46}$$

where $\omega \in [0,1]$ is a weighting coefficient between two decision modes, which represents the level of incompatibility of I and AB (resp., B). The coefficient $typ(AB)$ (resp., $typ(B)$) is the value of typicality associated to AB given I, as defined by typ_{exp} in section 2. Note that, in the analyzed experimental data set, typicality and probability judgements were performed by two distinct groups of participants, which may lead to additional discrepancies between theoretical and empirical values of the prospects' probabilities.

4.4.2 Measurement: operational prospects probabilities

A choice option is formulated as a composite prospect similar to (28) of subsection 4.3.2

$$\begin{aligned} \pi_{(AB)_1 I} &= (AB)_1 \otimes \biguplus_{k=1,2} I_k, \\ (resp., \quad \pi_{B_1 I} &= B_1 \otimes \biguplus_{k=1,2} I_k \ ), \end{aligned} \tag{47}$$

with a corresponding prospect state (29)

$$\begin{aligned} |\pi_{(AB)_1 I}\rangle &= \gamma_1 |(AB)_1 I_1\rangle + \gamma_2 |(AB)_1 I_2\rangle \\ (resp., \quad |\pi_{B_1 I}\rangle &= \gamma_1 |B_1 I_1\rangle + \gamma_2 |B_1 I_2\rangle \ ), \end{aligned} \tag{48}$$

which generates a prospect operator (30) (resp., (31))

$$\begin{aligned} \pi_{(AB)_1 I} \to \hat{P}_{(AB)_1 I} &\equiv |\pi_{(AB)_1 I}\rangle \langle \pi_{(AB)_1 I}| \\ &= \sum_k |\gamma_k|^2 |(AB)_1 I_k\rangle \langle (AB)_1 I_k| + \sum_{k \neq l} \gamma_k \gamma_l^* |(AB)_1 I_k\rangle \langle (AB)_1 I_l| \\ &= \sum_k |\gamma_k|^2 \hat{P}_{(AB)_1} \otimes \hat{P}_{I_k} + \sum_{k \neq l} \gamma_k \gamma_l^* \hat{P}_{(AB)_1} \otimes |I_k\rangle \langle I_l| \qquad k,l \in \{1,2\} \end{aligned} \tag{49}$$

$$\begin{aligned} (resp., \quad \pi_{B_1 I} \to \hat{P}_{B_1 I} &\equiv |\pi_{B_1 I}\rangle \langle \pi_{B_1 I}| \\ &= \sum_k |\gamma_k|^2 |B_1 I_k\rangle \langle B_1 I_k| + \sum_{k \neq l} \gamma_k \gamma_l^* |B_1 I_k\rangle \langle B_1 I_l| \\ &= \sum_k |\gamma_k|^2 \hat{P}_{B_1} \otimes \hat{P}_{I_k} + \sum_{k \neq l} \gamma_k \gamma_l^* \hat{P}_{B_1} \otimes |I_k\rangle \langle I_l| \qquad k,l \in \{1,2\} \ ). \end{aligned} \tag{50}$$

The uncertainty related to the description of a subject I, which is captured by an uncertainty union, again results in interfering modes $|I_k\rangle \langle I_l|$ $(k,l \in \{1,2\})$.

The probability that a decision maker in a state of mind $|\psi_{ABI}\rangle$ (resp., $|\psi_{BI}\rangle$) (45) will choose a prospect $\pi_{(AB)_1 I}$ (resp., $\pi_{B_1 I}$) is given by

$$\begin{aligned} p(\pi_{(AB)_1 I}) = \langle \psi_{ABI}| \hat{P}_{(AB)_1 I} |\psi_{ABI}\rangle &= |\langle \pi_{(AB)_1 I}|\psi_{ABI}\rangle|^2 \\ (resp., \quad p(\pi_{B_1 I}) = \langle \psi_{BI}| \hat{P}_{B_1 I} |\psi_{ABI}\rangle &= |\langle \pi_{B_1 I}|\psi_{BI}\rangle|^2) \end{aligned} \tag{51}$$

In an explicit form, we have

$$\begin{aligned} p(\pi_{(AB)_1 I}) &= |\gamma_1\zeta_{11}|^2 + |\gamma_2\zeta_{12}|^2 + 2Re(\zeta_{11}^*\gamma_1\gamma_2^*\zeta_{12}) \\ (resp.,\ \ p(\pi_{B_1 I}) &= |\gamma_1\kappa_{11}|^2 + |\gamma_2\kappa_{12}|^2 + 2Re(\kappa_{11}^*\gamma_1\gamma_2^*\kappa_{12})\ ), \end{aligned} \tag{52}$$

and it is comprised of two parts: (i) the diagonal term, i.e. utility factor:

$$\begin{aligned} f(\pi_{(AB)_1 I}) &= f(\pi_{(AB)_1 I_1}) + f(\pi_{(AB)_1 I_2}) = |\gamma_1\zeta_{11}|^2 + |\gamma_2\zeta_{12}|^2 \\ (resp.,\ \ f(\pi_{B_1 I}) &= f(\pi_{B_1 I_1}) + f(\pi_{B_1 I_2}) = |\gamma_1\kappa_{11}|^2 + |\gamma_2\kappa_{12}|^2\ ), \end{aligned} \tag{53}$$

and (ii) the off-diagonal term, i.e. attraction factor:

$$\begin{aligned} q(\pi_{(AB)_1 I}) &= 2Re(\zeta_{11}^*\gamma_1\gamma_2^*\zeta_{12}) \\ (resp.,\ \ q(\pi_{B_1 I}) &= 2Re(\kappa_{11}^*\gamma_1\gamma_2^*\kappa_{12})\ ). \end{aligned} \tag{54}$$

The uncertainty angle can be defined as (Yukalov and Sornette, 2009)

$$\begin{aligned} \Delta(\pi_{(AB)_1 I}) &\equiv \arg(\zeta_{11}^*\gamma_1\gamma_2^*\zeta_{12}) \\ (resp.,\ \ \Delta(\pi_{B_1 I}) &\equiv \arg(\kappa_{11}^*\gamma_1\gamma_2^*\kappa_{12})\ ). \end{aligned} \tag{55}$$

The uncertainty angle quantifies the “wedge” between the interfering decision modes of a prospect in consideration. In the proposed formulation, for a conjunctive category, the interfering modes are $|(AB)_1 I_1\rangle$ and $|(AB)_1 I_2\rangle$ and, for its constituent category, they are $|B_1 I_1\rangle$ and $|B_1 I_2\rangle$.

The introduction of the uncertainty angle (55) allows us to rewrite the attraction factor (54) as

$$\begin{aligned} q(\pi_{(AB)_1 I}) &= 2\sqrt{f(\pi_{(AB)_1 I_1}) f(\pi_{(AB)_1 I_2})}\ \cos\Delta(\pi_{(AB)_1 I}) = 2\sqrt{|\gamma_1\zeta_{11}|^2|\gamma_2\zeta_{12}|^2}\ \cos\Delta(\pi_{(AB)_1 I}) \\ (resp.,\ \ q(\pi_{B_1 I}) &= 2\sqrt{f(\pi_{B_1 I_1}) f(\pi_{B_1 I_2})}\ \cos\Delta(\pi_{B_1 I}) = 2\sqrt{|\gamma_1\kappa_{11}|^2|\gamma_2\kappa_{12}|^2}\ \cos\Delta(\pi_{B_1 I})\ ). \end{aligned} \tag{56}$$

To summarize, we propose that the judged probability p_{exp} that an instance I belongs to a conjunction category (resp., constituent category) is equal to the probability of a prospect $\pi_{(AB)_1 I}$ (resp., $\pi_{B_1 I}$). According to equations (52) and (56), we have

$$\begin{aligned} p_{exp}(AB)\ = p\ (\pi_{(AB)_1 I}) &= |\gamma_1\zeta_{11}|^2 + |\gamma_2\zeta_{12}|^2 + 2\sqrt{|\gamma_1\zeta_{11}|^2|\gamma_2\zeta_{12}|^2}\ \cos\Delta(\pi_{(AB)_1 I}) \\ (resp.,\ \ p_{exp}(B)\ = \ p(\pi_{B_1 I}) &= |\gamma_1\kappa_{11}|^2 + |\gamma_2\kappa_{12}|^2 + 2\sqrt{|\gamma_1\kappa_{11}|^2|\gamma_2\kappa_{12}|^2}\ \cos\Delta(\pi_{B_1 I})\ ). \end{aligned} \tag{57}$$

Note that one of the modes of a decision maker’s state (44) that represents an incompatibility of I and AB (resp., B), i.e. $|(AB)_2 I_1\rangle$ (resp., $|B_2 I_1\rangle$), is absent in the choice prospect (29) and, consequently, in the probability formulation of that prospect (57). Thus, in the focused mind state, when a decision maker is facing only one choice prospect, an incompatibility feature is (mostly) affecting a decision via the decision mode $|(AB)_1 I_2\rangle$ (resp., $|B_1 I_2\rangle$). We account for this by assigning a larger weight to $|(AB)_1 I_2\rangle$ (resp., $|B_1 I_2\rangle$), i.e. $\omega \in [0.5, 1]$ in (46). Thus, the coefficients of (57) are parameterised by

$$\begin{cases} |\zeta_{11}|^2 = typ(AB) \\ |\zeta_{12}|^2 = \omega(1 - typ(AB)) \\ \omega \in [0.5, 1] \end{cases} \left(\text{resp.,} \begin{cases} |\kappa_{11}|^2 = typ(B) \\ |\kappa_{12}|^2 = \omega(1 - typ(B)) \\ \omega \in [0.5, 1] \end{cases} \right) \tag{58}$$

The weight coefficient ω directly impacts the amplitudes of the decision modes. It also influences the uncertainty angle between them via the normalisation condition (40), since a change of amplitude of one component of a normalized vector amounts to a rotation.

The prospect probabilities, i.e. the probability judgements $p_{exp}(AB)$ and $p_{exp}(B)$ (57), are by definition real positive numbers $\mathfrak{R}_{\geq 0}$ in the interval $[0,1]$, and for complementary events sum to 1. Given the orthonormality of the basis (43), the squared probability amplitudes should also belong to $\mathfrak{R}_{\geq 0}$ in the interval $[0,1]$ and sum to 1 within a state of mind, or within a prospect. In the suggested analytical approach, for the coefficients $|\zeta_{ij}|^2$ and $|\kappa_{ij}|^2$, $i,j \in \{1,2\}$ of the state of mind, this condition is respected by their direct link to typicality judgements $typ(AB)$ and $typ(B)$, which were made by participants on the same scale from 0 to 1. In the subsequent analysis, special care should be taken to ensure that the coefficients $|\gamma_i|^2$, $i,j \in \{1,2\}$ of a prospect, as well as their sum, belong to $\mathfrak{R}_{\geq 0}$ in the interval $[0,1]$.

5 Empirical analysis and explanation of the conjunction fallacy

In this section, we apply the formulation of section (4.4) to the experimental results reported in (Shafir et al., 1990). We exploit equations (57) and (58) from two perspectives by (1) quantifying the uncertainty factor ($\cos\Delta(\pi)$) and (2) estimating the relative influence of interfering decision modes (coefficients $|\gamma_i|^2$).

5.1 First interpretation of the conjunction fallacy: Quantifying the uncertainty factor and uncertainty angle

Here, we focus on quantifying the uncertainty factor $\cos\Delta(\pi)$ associated with different types of categories: $AB(i)$, $B(i)$, $AB(c)$, $B(c)$. To allow for a meaningful comparison with experiments, we need to reduce the number of degrees of freedom contained in equations (57) and (58). We note that the coefficients γ_i, $i \in \{1,2\}$ embody the uncertainty of an instance I, which can be related to the briefness of the description of a subject and the imprecise time evolution of the described personal characteristics (captured by the uncertain union (9)). Thus, with no prior information, we can assume equal weights for attributing to a subject either characteristics directly mentioned in the description (i.e. I_1), or any other characteristics that are complementary to it (i.e. I_2). This amounts to imposing

$$|\gamma_1|^2 = |\gamma_2|^2 = 0.5 \; . \tag{59}$$

Taking into account equations (57)-(59), the attraction factor reduces to

$$\begin{aligned} q(\pi_{(AB)_1 I}) &= p_{exp}(AB) - 0.5\left[typ(AB) + \omega(1 - typ(AB))\right] \\ (resp., \quad q(\pi_{B_1 I}) &= p_{exp}(B) - 0.5\left[typ(B) + \omega(1 - typ(B))\right] \;), \end{aligned} \tag{60}$$

and the uncertainty factor reads

$$\begin{gathered} \cos\Delta(\pi_{(AB)_1 I}) = \frac{p_{exp}(AB) - 0.5\left[typ(AB) + \omega(1 - typ(AB))\right]}{\sqrt{typ(AB)\;\omega(1 - typ(AB))}} \\ \left(\text{resp.,} \quad \cos\Delta(\pi_{B_1 I}) = \frac{p_{exp}(B) - 0.5\left[typ(B) + \omega(1 - typ(B))\right]}{\sqrt{typ(B)\;\omega(1 - typ(B))}}\right). \end{gathered} \tag{61}$$

Based on the experimental results aggregated over all participants, i.e. the average probability judgement (p_{exp}) of an instance I to belong to one of the four categories and corresponding typicality judgements (typ_{exp}), the attraction factor q is presented in table 2 for each prospect. The corresponding uncertainty factors, i.e. cosines of the uncertainty angles, are shown in table 3. Both tables include results for the two boundary values $\omega = 1$ and $\omega = 0.5$. Values of $\cos\Delta(\pi)$ are constrained by the admissible range of the cosine function $[-1, +1]$. As shown in table 3, all but one (for $\omega = 1$: instance 12, $B(i)$) cosine values are found within this range. This indirectly supports the proposed analytical formulation.

Table 2: Attraction factor $q(\pi)$ by category aggregated over all participants over the 14 instances presented in (Shafir et al., 1990). The four categories are: an incompatible conjunction $AB(i)$ and its constituent $B(i)$, a compatible conjunction $AB(c)$ and its constituent $B(c)$. Sample mean μ and sample standard deviation σ are provided.

Instance I	Attraction factor $q(\pi)$:							
	for $\omega = 1$, for each category:				**for** $\omega = 0.5$, for each category:			
	$B(i)$	$AB(i)$	$B(c)$	$AB(c)$	$B(i)$	$AB(i)$	$B(c)$	$AB(c)$
1	-0.259	-0.099	0.033	0.101	-0.075	0.041	0.123	0.168
2	-0.327	-0.226	0.123	0.207	-0.108	-0.061	0.207	0.264
3	-0.340	-0.274	-0.145	-0.156	-0.134	-0.087	0.032	-0.017
4	-0.234	-0.133	0.098	0.186	-0.017	0.033	0.192	0.257
5	-0.298	-0.231	-0.016	0.052	-0.078	-0.053	0.063	0.120
6	-0.306	-0.218	0.183	0.050	-0.084	-0.030	0.264	0.149
7	-0.348	-0.123	0.044	0.078	-0.138	0.030	0.094	0.135
8	-0.312	-0.248	-0.093	0.016	-0.114	-0.094	0.057	0.125
9	-0.190	-0.029	-0.088	0.059	0.011	0.089	0.028	0.136
10	-0.328	-0.186	0.142	0.184	-0.114	-0.030	0.225	0.266
11	-0.185	0.080	-0.048	0.019	-0.006	0.190	0.067	0.082
12	-0.369	-0.251	0.207	0.009	-0.139	-0.086	0.262	0.115
13	-0.320	-0.161	-0.047	0.003	-0.099	0.016	0.075	0.086
14	-0.108	-0.061	0.129	0.133	0.085	0.083	0.212	0.213
μ	-0.280	-0.154	0.037	0.067	-0.072	0.003	0.136	0.150
σ	0.075	0.101	0.112	0.094	0.066	0.082	0.087	0.079

An important observation from tables 2 and 3 is that for both quantities - q and $\cos\Delta(\pi)$ - the absolute difference of their average values between the prospect with conjunction AB and the prospect with constituent B is large for incompatible (i) categories and is small for compatible (c) categories. This signals the higher dissimilarity within a (i) pair in comparison with a (c) pair pertaining to the same instance i. It holds for both values of $\omega \in \{1, 0.5\}$. The case of $\omega = 1$ seems to be more plausible, since it recovers closely the observed amplitudes of the conjunction fallacy and the conjunction effect, which were reported in table 1.

Thus, first, consider the case of $\omega = 1$. Table 3 shows that the absolute value of the uncertainty factor $|\cos\Delta(\pi)|$ is larger for the pairs of I with *incompatible* (i) categories. Noticeably, in all such pairs, the amplitude of both negative $\cos\Delta(\pi)$ and attraction factor q (table 2) are larger for a constituent $B(i)$ of an incompatible conjunction. All the categories $B(i)$ in the experiment were formulated such that the corresponding instance I would be atypical of that category (average $typ_{exp}(B(i)) = 0.167$). According to the experimental setup, the incompatibility of a conjunction category $AB(i)$ indicates that the included categories A and B share only a few common features, i.e. A and B are incompatible between each other. Thus, in the typicality and probability judgements concerning the conjunction $AB(i)$, the category A may partially compensate for the "repulsion" between I and $B(i)$, which is supported by the larger value of the average $typ_{exp}(AB(i)) = 0.371 > typ_{exp}(B(i))$.

Table 3: Uncertainty factor (cosine of the uncertainty angle $\cos\Delta(\pi)$) by category aggregated over all participants over the 14 instances presented in (Shafir et al., 1990). The four categories are: an incompatible conjunction $AB(i)$ and its constituent $B(i)$, a compatible conjunction $AB(c)$ and its constituent $B(c)$. Sample mean μ and sample standard deviation σ are provided. Values of $\cos\Delta(\pi)$ are constrained by the admissible range of the cosine function $[-1, +1]$; unadjusted values are in brackets.

Instance I	Cosine of uncertainty angle $\cos\Delta(\pi)$:							
	for $\omega = 1$, for each category:				**for** $\omega = 0.5$, for each category:			
	$B(i)$	$AB(i)$	$B(c)$	$AB(c)$	$B(i)$	$AB(i)$	$B(c)$	$AB(c)$
1	-0.588	-0.199	0.069	0.228	-0.241	0.118	0.362	0.536
2	-0.989	-0.477	0.260	0.494	-0.463	-0.182	0.620	0.890
3	-0.891	-0.631	-0.318	-0.314	-0.497	-0.283	0.098	-0.049
4	-0.689	-0.281	0.202	0.412	-0.072	0.098	0.561	0.806
5	-0.914	-0.510	-0.034	0.117	-0.339	-0.166	0.190	0.381
6	-0.978	-0.504	0.391	0.102	-0.377	-0.099	0.798	0.431
7	-0.947	-0.252	0.110	0.186	-0.532	0.086	0.333	0.455
8	-0.769	-0.510	-0.190	0.032	-0.397	-0.273	0.165	0.356
9	-0.480	-0.058	-0.177	0.128	0.040	0.253	0.078	0.416
10	-0.934	-0.384	0.302	0.392	-0.459	-0.087	0.676	0.801
11	-0.411	0.161	-0.096	0.044	-0.017	0.542	0.190	0.266
12	-1 (-1.376)	-0.530	0.499	0.018	-0.730	-0.257	0.894	0.330
13	-0.999	-0.354	-0.094	0.006	-0.437	0.051	0.212	0.258
14	-0.257	-0.123	0.274	0.285	0.285	0.238	0.636	0.646
μ	-0.775 (-0.801)	-0.332	0.086	0.152	-0.303	0.003	0.415	0.466
σ	0.249 (0.292)	0.221	0.245	0.206	0.272	0.240	0.275	0.254

This consequently leads to a higher, though still negative, attraction factor for $AB(i)$, compared to its constituent $B(i)$.

Continuing the case of $\omega = 1$, for compatible categories $AB(c)$ and $B(c)$, the uncertainty factors $\cos\Delta(\pi)$ are positive, but much smaller in amplitude than for incompatible pairs. Overall, for compatible categories, there is a slight positive attraction effect. For 12 (resp., 11) out of 14 instances, $\cos\Delta(\pi)$ (resp., q) is larger for a compatible conjunction $AB(c)$ than for a related constituent $B(c)$. Thus, the typicality of $B(c)$ (average $typ_{exp}(B(c)) = 0.606$) is enhanced by a compatible feature A in conjunction $AB(c)$ (average $typ_{exp}(AB(c)) = 0.670$), and increases the positive attraction and probability judgements for the latter.

A change in ω from 1 to 0.5 amounts to a rotation. As expected, for the smaller value of $\omega = 0.5$, the absolute value of the attraction factor q and of the uncertainty factor increases in comparison with $\omega = 1$. This means that, in order to explain empirical data, a higher uncertainty factor is needed to compensate for a smaller amplitude ζ_{12} (resp., κ_{12}) of the incompatible decision mode $|(AB)_1 I_2\rangle$ (resp., $|B_1 I_2\rangle$) in a decision makers' state of mind (see equation 58). The main distinction of the case with $\omega = 0.5$ is that the positive values of q and $\cos\Delta(\pi)$ for the compatible (c) categories become profound, and even increase in magnitude the corresponding values for incompatible (i) pairs.

Figure 2 shows the average over 14 instances I of the uncertainty angle $\Delta(\pi)$ for the four categories and their shift when ω changes from 1 (left subplot) to 0.5 (right subplot). This intuitive representation reveals the differences between incompatible and compatible categories with respect

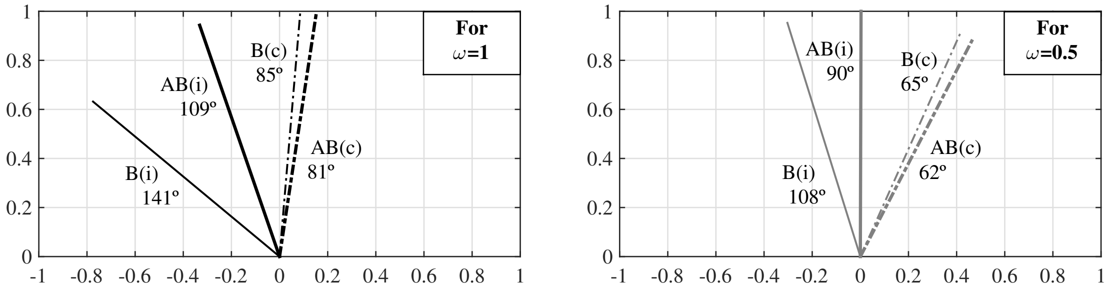

Figure 2: Average over 14 instances I of the uncertainty angle $\Delta(\pi)$ with $\omega = 1$ (left subplot, black) or $\omega = 0.5$ (right subplot, grey), for each of the four categories: an incompatible conjunction $AB(i)$ (bold solid line), and its constituent $B(i)$ (solid line), a compatible conjunction $AB(c)$ (bold dash-dotted line), and its constituent $B(c)$ (dash-dotted line). The conjunction fallacy is associated with the higher dissimilarity within the pair of incompatible (i) prospects. For the main considered case of $\omega = 1$ (left subplot), the uncertainty factor $\cos\Delta(\pi)$ (and resulting attraction factor q) is slightly positive for (c) categories, and profoundly negative for (i) categories. The average negative $\cos\Delta(\pi)$ is almost twice larger in amplitude for a constituent $B(i)$ compared with a conjunction $AB(i)$, suggesting an explanation of the conjunction fallacy.

to the associated uncertainty factors. In particular, for the case of $\omega = 1$, $\cos\Delta(\pi)$ (and resulting attraction factor q) is slightly positive for compatible (c) categories, and profoundly negative for incompatible (i) ones. Herewith, the average negative $\cos\Delta(\pi)$ is almost twice larger for a constituent $B(i)$ compared with a conjunction $AB(i)$. This leads to a high unattractiveness of the latter, and low probability (judgement) of the corresponding prospect. This constitutes one of the plausible perspectives to interpret the conjunction fallacy. For the case of $\omega = 0.5$, the absolute values of $\cos\Delta(\pi)$ increase, but the same mechanism (i.e. higher dissimilarity between incompatible (i) prospects) explains the observed conjunction fallacy.

5.2 Second interpretation of the conjunction fallacy: Estimating the relative influence of interfering decision modes ($|\gamma_1|^2$ and $|\gamma_2|^2$)

We now analyze equations (57) from a different perspective, namely as a system of two coupled equations. The analysis is conducted for each of the 28 triples defined as follows. A compatible triple (I, $AB(c)$, $B(c)$) consists in an instance I combined with two compatible categories $AB(c)$ and $B(c)$. An incompatible triple (I, $AB(i)$, $B(i)$) consists in an instance I combined with two incompatible categories $AB(i)$ and $B(i)$. Using the 14 instances I studied in (Shafir et al., 1990), we thus have 28 triples to consider.

Recall that the uncertainty, which originates from the indeterminacy of a subject's description I, was introduced as an uncertain union (9). We assume that this uncertainty has a similar influence onto the two prospects that are associated with the categories (AB and B) of the same type (compatible or incompatible). This assumption amounts to imposing that the coefficients γ_1 and γ_2 are the same within each triple (I, AB, B).

Again, to allow for a meaningful comparison with experiments, we need to reduce the number of degrees of freedom contained in equations (57) and (58). We thus consider two extreme cases: minimum (vanishing) interference and maximum interference.

- Minimum interference ($q_{min} = 0$) is achieved for $\cos\Delta(\pi) = 0$, i.e. when the interfering decision modes are orthogonal ($\Delta(\pi) = 90°$). Solving simultaneously both equations from (57), we obtain:

$$\begin{cases} |\gamma_1|^2 = \dfrac{p_{exp}(AB) - |\gamma_2|^2|\zeta_{12}|^2}{|\zeta_{11}|^2} \\ |\gamma_2|^2 = \dfrac{p_{exp}(B)|\zeta_{11}|^2 - p_{exp}(AB)|\kappa_{11}|^2}{|\zeta_{11}|^2|\kappa_{12}|^2 - |\zeta_{12}|^2|\kappa_{11}|^2} \end{cases}$$

Taking into account (58), this leads to

$$\begin{cases} |\gamma_1|^2 = \dfrac{p_{exp}(AB) - |\gamma_2|^2\omega(1 - typ(AB))}{typ(AB)} \\ |\gamma_2|^2 = \dfrac{1}{\omega}\dfrac{p_{exp}(B)typ(AB) - p_{exp}(AB)typ(B)}{typ(AB) - typ(B)} \end{cases} \tag{62}$$

- Maximum interference occurs when $|\cos\Delta(\pi)| = 1$, i.e. when the interfering decision modes are collinear. For maximum positive (resp., negative) interference, when the uncertainty angle between interfering decision modes $\Delta(\pi) = 0^\circ$ and $\cos\Delta(\pi) = 1$ (resp., $\Delta(\pi) = 180^\circ$ and $\cos\Delta(\pi) = -1$), equations (57) give for $|\gamma^+|$ (resp., $|\gamma^-|$):

  $$\begin{cases} |\gamma_1^\pm| = \dfrac{\sqrt{p_{exp}(AB)} \mp |\gamma_2^\pm||\zeta_{12}|}{|\zeta_{11}|} \\ |\gamma_2^\pm| = \pm\dfrac{\sqrt{p_{exp}(B)}|\zeta_{11}| - \sqrt{p_{exp}(AB)}|\kappa_{11}|}{|\zeta_{11}||\kappa_{12}| - |\zeta_{12}||\kappa_{11}|} \end{cases}$$

  Taking into account (58), this leads to

  $$\begin{cases} |\gamma_1^\pm| = \dfrac{\sqrt{p_{exp}(AB)} \mp |\gamma_2^\pm|\sqrt{\omega}\sqrt{(1 - typ(AB))}}{\sqrt{typ(AB)}} \\ |\gamma_2^\pm| = \pm\dfrac{1}{\sqrt{\omega}}\dfrac{\sqrt{p_{exp}(B)typ(AB)} - \sqrt{p_{exp}(AB)typ(B)}}{\sqrt{typ(AB)(1 - typ(B))} - \sqrt{(1 - typ(AB))typ(B)}} \end{cases} \tag{63}$$

  Note that $|\gamma_2^+| = -|\gamma_2^-|$ and $|\gamma_1^+| = |\gamma_1^-|$. Interestingly, in the expressions of $|\gamma_1|$, $\sqrt{\omega}$ cancels out when the value of $|\gamma_2|$ is substituted into it, implying that $|\gamma_1|$ does not depend on ω. Moreover, for positive and negative maximum interferences, the squared coefficients are equal: $|\gamma_2^+|^2 = |\gamma_2^-|^2$ and $|\gamma_1^+|^2 = |\gamma_1^-|^2$. These squared coefficients are the mean objects of interest since they are interpreted as the weights of the corresponding decision modes of a prospect.

For all 28 triples (I, AB, B), the squared coefficients $\{|\gamma_i|^2, i \in \{1,2\}\}$ were calculated for the cases of both minimum (zero) and maximum interference, and for the two values $\omega = 1$ and $\omega = 0.5$. The values of the squared probability amplitudes $|\gamma_i|^2$, $i \in \{1,2\}$, and their sums, for 28 triples (I,AB,B) and all considered conditions (minimum and maximum interference, and $\omega = \{1, 0.5\}$), are reported in appendix A.3 (tables 9 and 10). Raw values, prior to the adjustments mentioned at the end of this subsection (concerning small deviations from the allowed window and outliers), are provided in brackets.

Figure 3 presents the complementary cumulative distribution function (CCDF) of these coefficients and their sum over the set of the 28 triples. Note the predominance of coefficient $|\gamma_1|^2$, which characterizes the compatibility (typicality) between a described subject (I) and categories in a triple (AB and B). Thus, $|\gamma_1|^2$ is the major contribution within a prospect that influences the probability for that prospect to be chosen (i.e. probability judgment). Coefficient $|\gamma_2|^2$, which is associated with the incompatibility (atypicality) between a considered subject (I) and the triple's categories (AB and B), plays a secondary role. It is inversely proportional to $\omega \in \{1, 0.5\}$, as was shown analytically in equations (63). This dependence has a simple explanation. *Within an entangled state of mind* (44), if the atypicality factor is not loaded purely onto the mode $|AB_1I_2\rangle$ ($|B_1I_2\rangle$), but is alos distributed onto an additional mode $|AB_2I_1\rangle$ ($|B_2I_1\rangle$) via decreasing ω from 1 to 0.5, then the observed prospect probabilities can be explained only by a proportional increase of the influence of the original atypicality modes $|AB_1I_2\rangle$ and $|B_1I_2\rangle$ *in the prospect.*

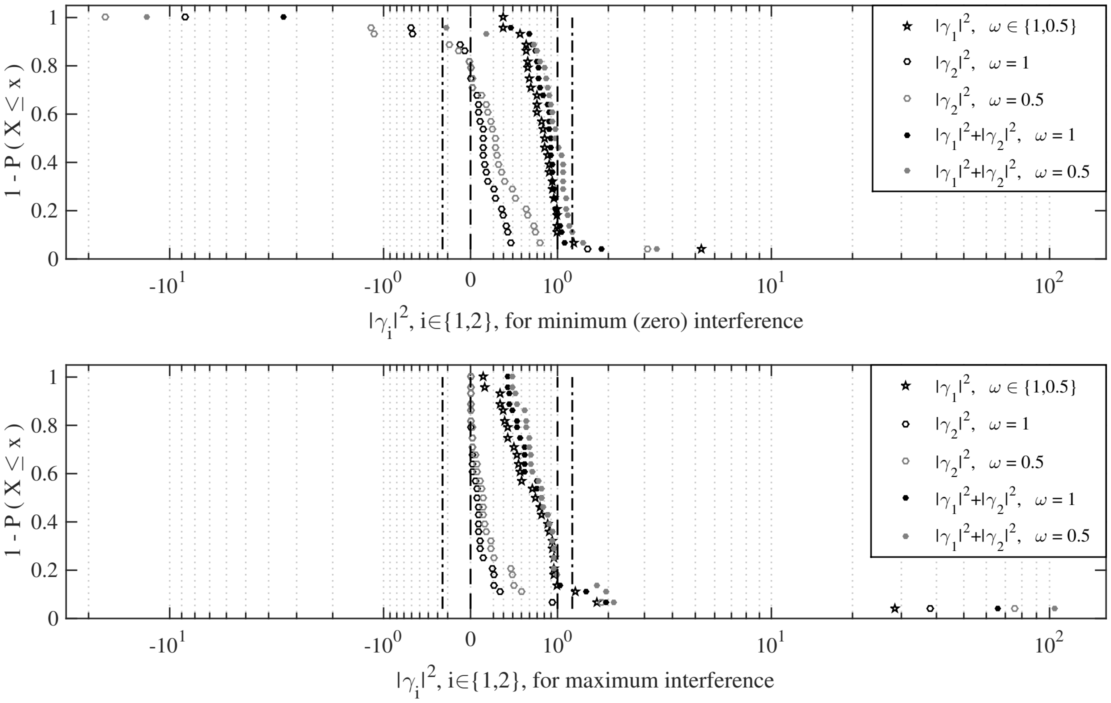

Figure 3: Complementary cumulative distribution function (CCDF) of squared probability amplitudes $\{|\gamma_i|^2, i \in \{1,2\}\}$, for minimum (upper panel) and maximum interference (lower panel) cases, over the set of the 28 triples described in the text. The coefficients $\{|\gamma_i|^2, i \in \{1,2\}\}$ are the probability weights of interfering decision modes within a prospect, and are calculated for all 28 triples (I,AB,B), including 14 incompatible (i) and 14 compatible (c). Note the predominance of coefficient $|\gamma_1|^2$ (star), which is associated with the typicality between a described subject (I) and two categories in the triple (AB and B). Coefficient $|\gamma_2|^2$ (empty circle), which quantifies the atypicality between a considered subject (I) and the triple's categories (AB and B), is inversely proportional to $\omega \in \{1, 0.5\}$, i.e. to the weight of the prospect's atypicality decision mode in the entangled state of mind (44). Most of the observed values of coefficients $|\gamma_i|^2$ and their sum $|\gamma_1|^2 + |\gamma_2|^2$ (filled black circle for $\omega = 1$ and filled grey circle for $\omega = 0.5$) are within their allowed region $[0, 1]$ (dashed vertical lines), in particular for the maximum interference case. The extended region $[-0.25, 1.25]$ (dash-dotted vertical lines) captures values in the vicinity of the allowed region, especially relevant for the case of minimum interference. A logarithmic transformation that is symmetric with respect to the origin on the x-axis has been performed with the Matlab *symlog* function, which was created by R. Perrotta, and based on (Webber, 2013).

Figure 3 shows that, for each triple, the calculated values of $|\gamma_i|^2$ and of their sum $|\gamma_1|^2 + |\gamma_2|^2$ are found mostly inside their allowed region $[0, 1]$. Thus, the condition that constrains the squared probability amplitudes of a prospect and their sum to be real positive numbers $\mathfrak{R}_{\geq 0}$ in the interval $[0, 1]$, which was introduced at the end of section 4.4, is in general satisfied. This is especially true for the maximum interference case (lower subplot of figure 3), as well as for the minimum interference case with $\omega = 1$. For the minimum interference case (upper subplot), in particular for

$\omega = 0.5$, most of the values that fall outside the allowed interval $[0,1]$ remain close to it. Taken together, the above observations support overall the proposed analytical approach as applied to the experiments of (Shafir et al., 1990).

Concerning the cases that lead to failures to fall in the allowed interval $[0,1]$, two groups should be distinguished.

- Small departures from $[0,1]$ can be considered insignificant, because they could arise from minor inconsequencial causes, such as the unavoidable simplifications and approximations in the parameterisation of the theoretical framework, imperfections in the experimental design, erroneous answers ("noise") of decision makers, and so on. Thus, in the subsequent analysis, these small deviations are adjusted to satisfy the constraint $\mathfrak{R}_{\geq 0} \in [0,1]$, i.e. slightly negative values are replaced by 0, and when $|\gamma_1|^2 + |\gamma_2|^2 > 1$, we normalise $|\gamma_1|^2$ and $|\gamma_2|^2$ correspondingly so that the sum becomes 1.
- Coefficient and/or sum values that are far away from the allowed region are considered "outliers." They pose a challenge to our theoretical framework. They should not be rejected by analysed carefully as their origin may point to new insights. This is the purpose of the next subsection.

5.3 Outliers detection and analysis

Several approaches can be followed to identify outliers. The simple visual analysis (heuristic approach) of figure 3 suggests extending the allowed region from $[0,1]$ to $[-0.25, 1.25]$ to account for noise and imperfections in the experiments and in the theoretical parameterisation. This extended allowed region contains most of the values that are outside $[0,1]$, even for the case of minimum interference with $\omega = 0.5$.

Another approach consists in using a robust statistical approach based on the interquartile range (IQR). In this way, boundaries to identify outliers could be determined by a multiple M of the IQR, which is subtracted from (resp., added to) the first quartile Q_1 (resp., third quartile Q_3). A desired property of these boundaries would be that they cover and enlarge the allowed range of the data. However, due to the particularities of the analyzed data set, this desired property cannot be achieved with the same M for all types of coefficients. In particular, the distributions of $|\gamma_2|^2$ are centered closer to 0, while the distributions of $|\gamma_1|^2$ and $|\gamma_1|^2 + |\gamma_2|^2$ gravitate towards 1. This leads to unintuitive boundaries that are skewed to the negative side for $|\gamma_2|^2$, and at the same time skewed to the right side (above 1) for other coefficients. Despite this shortcoming, we use the IQR-based method to determine median value of multiplier M, among all types of γ coefficients, which is implied by the extended allowed region $[-0.25, 1.25]$ introduced from the visual analysis of figure 3. The implied median $M = 1.5$, which is a widely used multiplier's value with the IQR method, applied to normally distributed datasets. This finding indirectly supports the extended allowed region $[-0.25, 1.25]$ of the heuristic approach.

A better grounded approach to identify outliers requires formal statistical testing. For this, an approximate theoretical distribution of the coefficients γ should be determined. Quantile-quantile (Q-Q) plots allow for graphical comparison of empirical probability distributions of the calculated raw coefficients with the theoretical normal probability distribution (figure 4). For all considered cases (minimum and maximum interference, and $\omega \in \{1, 0.5\}$), the Q-Q plots support the normality assumption for the coefficients $\{|\gamma_i|^2, i \in \{1,2\}\}$ and their sums, and moreover expose potential outliers.

As a formal statistical test, a generalized (extreme Studentized deviate) ESD test is applied, which

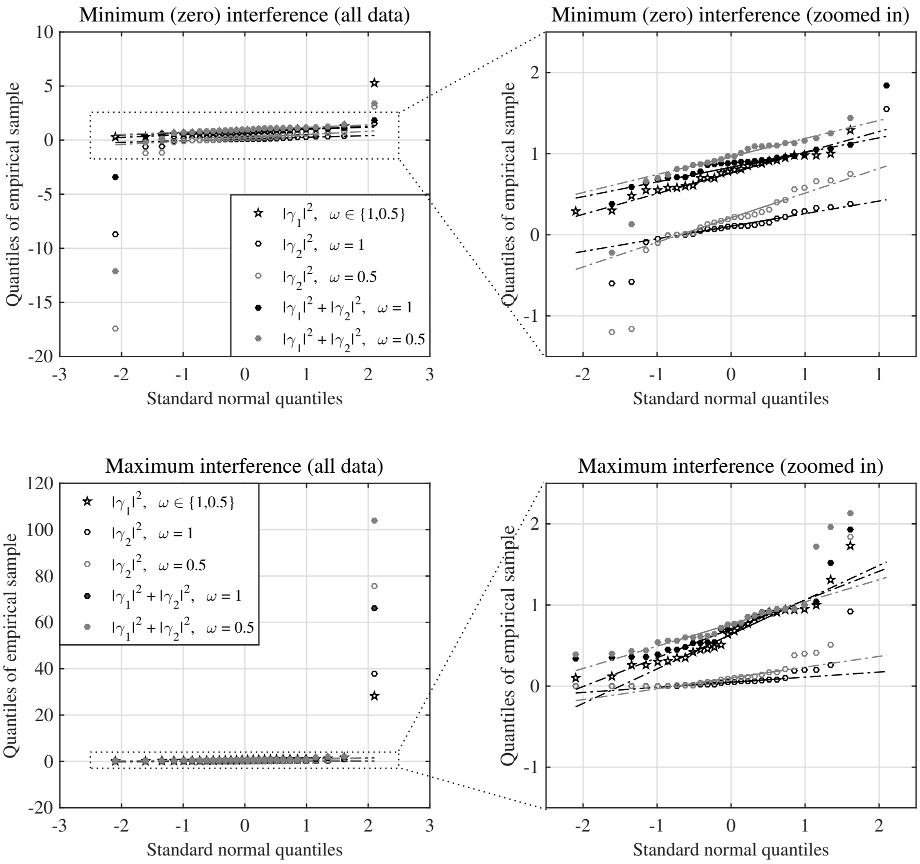

Figure 4: Quantile-quantile (Q-Q) plots that compare empirical probability distributions of the calculated row coefficients (i.e squared probability amplitudes) $|\gamma_1|^2$ and $|\gamma_2|^2$ and their sums, to the theoretical normal probability distribution. For both - minimum (zero) and maximum interference cases (resp., upper and lower subplots), and both $\omega \in \{1, 0.5\}$, the Q-Q plots support the normality assumption for coefficients $|\gamma_1|^2$ and $|\gamma_2|^2$ and their sums. Moreover, potential outliers are exposed. Plots on the left side present the whole datasets, while plots on the right side zoom them in, for better appreciation of central part of the distributions

is a many-outlier detection procedure. It is suitable for the identification of one or more outliers in a univariate dataset that follows an approximately normal distribution. The generalized ESD test was shown to be adequately accurate for detecting up to 10 outliers in samples as small as 25 (Rosner, 1983). To detect outliers, the generalized ESD procedure was repeated separately - for each coefficient γ_i, $i \in \{1, 2\}$, and their sum, and under each condition (minimum and maximum interference, and $\omega \in \{1, 0.5\}$). The results of the two-sided test with significance level $\alpha^{GESD} = 0.001$ are reported in Appendix A.3, table 11. In total, four outliers were identified: 10 (c), 7 (c), 5(c) and 6 (c), where names correspond to an index number of an instance I and a type of categories AB, B - compatible (c) or incompatible (i). The same outliers appear under all considered conditions of interference and values of ω, though the number of outliers differs. One condition, maximum interference with $\omega = 1$ stands out, as it includes the least number of outliers (two), which provides additional support to the validity of this parameterisation.

The identified outliers have been marked by an asterisk in tables 9 and 10 of Appendix A.3. If, for a given triple $(I,\ AB,\ B)$, at least one of the coefficients $|\gamma_i|^2, i \in \{1,2\}$, or their sum, is identify as an outlier, than all triple's coefficients for that condition are marked also as outliers, and excluded from the general dataset for further calculations. However, outliers should not be thrown away. Their separate analysis can provide useful insights about the origins of the abnormal values of the coefficients and inform on the limits of the proposed analytical approach.

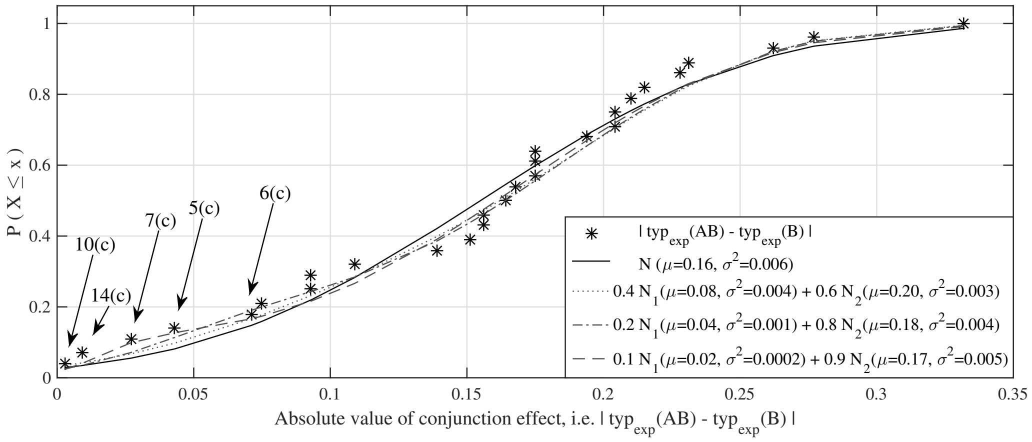

Figure 5: Cumulative distribution function (CDF) of an absolute value of conjunction effect, i.e. $|typ_{exp}(AB) - typ_{exp}(B)|$, best fitted by a single Gaussian model. All the outliers, which were identified based on prospects' squared probability amplitudes $|\gamma_i|^2$ $(i \in \{1,2\})$ and their sums, are marked by arrows and concentrated at the far left tail, where conjunction effect → 0. The only exception - 14 (c) - is not an outlier, but its corresponding γ coefficients exceeded the allowed region and were subject to adjustment, thus confirming specific nature of the left tail. We associate this effect with limit of discriminating ability (of a decision maker, or an experimental procedure), and term it *"QDT indeterminacy (uncertainty) principle"*

We have identified a number of outliers based on the above analysis of the probability amplitudes $|\gamma_1|^2$ and $|\gamma_2|^2$ and their sums. Are there also outliers directly observable in the distribution of the empirical conjunction effect? To address this question, figure 5 shows the distribution of the absolute values $|typ_{exp}(AB) - typ_{exp}(B)|$ of the conjunction effect, which can be well approximated by a normal probability distribution $N(\mu = 0.1550, \sigma = 0.0064)$. For this sample, no outliers were identified with a formal generalized ESD test. Note that most of the tail values, both left and right side, lay above the single Gaussian model (solid line), while values in the center of the sample are found below the theoretical line. These systematic deviations motivate the search for an improved approximation model. Thus, the single Gaussian (G) model was compared with Gaussian mixture (GM) models. Several mixtures with different proportions of constituent normal distributions are drawn in figure 5, and results of their multiple-criteria testing against a single Gaussian model is presented in table 4. Though the log-likelihood objective function is found maximum (or minimum for the negative log-likelihood) under the (unrestricted) Gaussian mixture model assumptions, this improvement of the fit is not sufficient to compensate for the cost associated with the three additional parameters. The single (restricted) Gaussian model performs better for all criteria of quality of fit, such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). According to the AIC of the best performing mixture model (GM: $0.1N_1 + 0.9N_2$) with respect to a single Gaussian model, the former is only 0.28 times as probable as the Gaussian model to minimize the information loss. In addition, the log-likelihood ratio test of nested hypothesis (Wilks, 1938) does not reject the single Gaussian model (p-value = 0.33).

Table 4: Selection of the best fitting model for the empirical distribution of the absolute value of the conjunction effect, i.e. $|typ_{exp}(AB) - typ_{exp}(B)|$. A single Gaussian (G) model is tested against several Gaussian mixture (GM) models, with different proportions of constituent normal distributions, and is found to perform best. Figure 5 illustrates the fits.

Model selection:		Negative log-likelihood	AIC	Relative likelihood, $e^{\frac{AIC_G - AIC_{GM}}{2}}$	BIC	Log-likelihood ratio test for nested models, p-value
Single Gaussian (G) model	N	-30.92	**-57.85**	**1.00**	**-55.18**	+
Gaussian mixture (GM) model:	$0.1N_1+0.9N_2$	**-32.65**	-55.30	0.28	-48.64	0.33
	$0.2N_1+0.8N_2$	-32.14	-54.28	0.17	-47.61	0.49
	$0.4N_1+0.6N_2$	-31.37	-52.75	0.08	-46.09	0.83

Thus, on the one hand, we demonstrate clearly the existence of outliers in the set of probability amplitudes $|\gamma_1|^2$ and $|\gamma_2|^2$ and their sums, while no outlier exists in the set of conjunction effect amplitudes. This apparent contradiction is resolved by the following observation that sheds light on a potential reason of the appearance of the former outliers. In figure 5, which presents the cumulative distribution function (CDF) of the absolute value $|typ_{exp}(AB) - typ_{exp}(B)|$ of the conjunction effect, one can observe that all outliers identified by the above analysis of the probability amplitudes $|\gamma_1|^2$ and $|\gamma_2|^2$ and their sums are found to be concentrated at the far left tail of the distribution, and are associated with the lowest magnitudes of the observed conjunction effect. This means that the abnormal values arise in the situation when typicality judgements of the two categories in a triple (AB and B) are very close to each other. On the one hand, this can be understood analytically from equations (62) and (63). For instance, in equations (62), $|\gamma_2|^2$ shots up dramatically as the difference in typicality judgements becomes very small: $(typ_{exp}(AB) - typ_{exp}(B)) \to 0 \implies |\gamma_2|^2 \to \infty$. At first sight, this property could be considered as a limitation of the proposed analytical formulation. On the other hand, if the theoretical formulation is adequate for most of the choice situations, its inapplicability to certain prospects (judgments) with extremely small differentiating characteristics may reveal a limitation of a more fundamental nature: a ceiling in discriminating abilities, either of a decision maker, or of an experimental procedure. Some differences may simply be *too small to notice*. This can be termed the *"QDT indeterminacy (uncertainty) principle"*, as representing a fundamental limit to the precision with which certain pairs (sets) of prospects can be simultaneously known (assessed) by a decision maker, or elicited by an experimental procedure. The formulation of the "QDT indeterminacy (uncertainty) principle", and its characterization, e.g. as a special regime, is proposed as an important and promising research direction. However, larger empirical datasets under various conditions should be analyzed before arriving to a conclusive understanding.

Note that, in the proposed analytical formulation, typicality judgements (typ) are directly linked to the weights ζ and κ of decision modes *in the state of mind* of a decision maker (see equations (58)). In turn, the coefficients γ are associated to the uncertainty of interfering decision modes *within a prospect*, which is captured by an uncertain union (see equation (9)). The interconnection between the two sources of indeterminacy (uncertainty), and towards the considered "QDT indeterminacy (uncertainty) principle", calls for further investigation. However, the current analysis provides some first evidence for this novel QDT proposition.

5.4 Return to the second interpretation of the conjunction fallacy: Analysis of the adjusted coefficients $|\gamma_1|^2$ and $|\gamma_2|^2$

As a result of the detailed analysis of the coefficients γ, their raw values were adjusted as explained in the previous subsection: small deviations were corrected to satisfy the constraint $\Re_{\geq 0} \in [0, 1]$, and outliers were excluded from further calculations. Adjusted (and raw) values are reported in appendix A.3 (tables 9 and 10). The adjusted coefficients γ aggregated over the triples (I, AB, B) of different types (i) or (c) are included in table 5.

Table 5: Aggregated adjusted squared probability amplitudes $|\gamma_i|^2$, $i \in \{1, 2\}$, and their sums for the minimum (zero) interference and maximum interference, as defined in subsection 5.2. The sample mean μ and sample standard deviation σ for triples (I, AB, B) of different types - incompatible (i) and compatible (c) - are reported. The coefficients $|\gamma_1|^2$ and $|\gamma_2|^2$ represent the weights of the interfering decision modes $|(AB)_1 I_1\rangle$ and $|(AB)_1 I_2\rangle$ within a prospect, i.e. the probability judgement of an instance I to belong to a category (AB or B)

	Minimum (zero) interference, $\cos\Delta(\pi) = 0$						Maximum interference, $\lvert\cos\Delta(\pi)\rvert = 1$					
For $\omega = 1$:	$\lvert\gamma_1\rvert^2$		$\lvert\gamma_2\rvert^2$		$\sum_{i=1,2} \lvert\gamma_i\rvert^2$		$\lvert\gamma_1\rvert^2$		$\lvert\gamma_2\rvert^2$		$\sum_{i=1,2} \lvert\gamma_i\rvert^2$	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
24 triples (I,AB,B), incl.:	**0.74**	0.18	0.13	0.12	**0.87**	0.12	**0.62**	0.28	0.05	0.06	**0.68**	0.24
14 triples $(I,AB(i),B(i))$	**0.73**	0.17	0.11	0.09	**0.84**	0.14	**0.55**	0.28	0.05	0.05	**0.60**	0.24
10 triples $(I,AB(c),B(c))$	**0.74**	0.20	0.16	0.14	**0.90**	0.09	**0.71**	0.26	0.06	0.08	**0.77**	0.21
For $\omega = 0.5$:	$\lvert\gamma_1\rvert^2$		$\lvert\gamma_2\rvert^2$		$\sum_{i=1,2} \lvert\gamma_i\rvert^2$		$\lvert\gamma_1\rvert^2$		$\lvert\gamma_2\rvert^2$		$\sum_{i=1,2} \lvert\gamma_i\rvert^2$	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
24 triples (I,AB,B), incl.:	**0.70**	0.19	0.23	0.20	**0.93**	0.09	**0.60**	0.28	0.10	0.13	**0.70**	0.19
14 triples $(I,AB(i),B(i))$	**0.71**	0.18	0.21	0.16	**0.91**	0.10	**0.55**	0.28	0.10	0.11	**0.65**	0.21
10 triples $(I,AB(c),B(c))$	**0.69**	0.22	0.28	0.25	**0.97**	0.05	**0.68**	0.27	0.10	0.16	**0.77**	0.14

The first notable observation from table 5 is that the sample means of the sum $|\gamma_1|^2 + |\gamma_2|^2$, for the two decision modes that make up a prospect, are close to 1, especially for the minimum interference case. Thus, for this case, the prospect states are found approximately normalized, which is a characteristic property of a decision-maker's state of mind. Within QDT, a decision consists in the transition of a current state of mind into a new state, which is equivalent to the chosen prospect. During the transition, the state vector of a chosen prospect is being normalized, so that it corresponds to the new state of mind. The results of our analysis show that the prospects in the deliberation phase, when interference is minimal, are already close to being normalized. In other words, even before the transition (decision) occurs, prospects resemble a normalized state of mind.

As expected from derivations (62)-(63), for both interference cases (minimum and maximum interference), changing ω from 1 to 0.5 increases the squared coefficients $|\gamma_2|^2$ by a factor of 2. As was previously mentioned, when the influence of the atypicality factor in the mind of a decision maker is redistributed evenly from one decision mode $|(AB)_1 I_2\rangle$ onto two modes $|(AB)_1 I_2\rangle$ and $|(AB)_2 I_1\rangle$, the squared probability amplitude of the former atypicality mode $|(AB)_1 I_2\rangle$ in the prospect is

Table 6: Results of two-sample t-tests for equal means (without assumption of equal variances) between two types of triples. H_0: means of $|\gamma_1|^2$ (or $|\gamma_1|^2 + |\gamma_2|^2$) for both types of triples $(I{,}AB(i){,}B(i))$ and $(I{,}AB(c){,}B(c))$ are *equal*; H_1: means of $|\gamma_1|^2$ (or $|\gamma_1|^2 + |\gamma_2|^2$) for triples $(I{,}AB(i){,}B(i))$ and $(I{,}AB(c){,}B(c))$ are *unequal*; H_1^* (a modified one-side hypothesis, reported in brackets): mean of $|\gamma_1|^2$ (or $|\gamma_1|^2 + |\gamma_2|^2$) for $(I{,}AB(i){,}B(i))$ *is less* than the mean for $(I{,}AB(c){,}B(c))$. Though reasonable significance levels are achieved only for $|\gamma_1|^2 + |\gamma_2|^2$, the general tendency is confirmed: in the case of minimum interference, the two types of triples - (i) and (c) - are indistinguishable, while for maximum interference (i) and (c) triples manifest themselves differently.

	Discriminability of triple's types - (i) and (c) - within each interference case:							
	$\lvert\gamma_1\rvert^2$				$\lvert\gamma_1\rvert^2 + \lvert\gamma_2\rvert^2$			
	Minimum (zero) interference		Maximum interference		Minimum (zero) interference		Maximum interference	
For $\omega = 1$:	μ	p-value	μ	p-value	μ	p-value	μ	p-value
$(I{,}AB(i){,}B(i))$	0.73	**0.907** **(0.454)**	0.55	**0.133** **(0.067)**	0.84	**0.253** **(0.127)**	0.60	**0.062** **(0.030)**
$(I{,}AB(c){,}B(c))$	0.74		0.71		0.90		0.77	
For $\omega = 0.5$:	μ	p-value	μ	p-value	μ	p-value	μ	p-value
$(I{,}AB(i){,}B(i))$	0.71	0.871 (0.565)	0.55	0.268 (0.134)	0.91	0.100 (0.050)	0.65	0.103 (0.051)
$(I{,}AB(c){,}B(c))$	0.69		0.68		0.97		0.77	

raised twofold in order to explain the empirical values of the prospects' probabilities (probability judgements).

The most interesting insight from table 5 is a manifestation of triples with distinct category types - incompatible (i) and compatible (c) - in the two different interference cases. For a moment, consider condition $\omega = 1$, which required smaller coefficient adjustments and, thus, can be perceived as a more reliable parameterisation. When zero interference between decision modes is imposed, the means of the squared coefficients $|\gamma_1|^2$, as well as the means of $|\gamma_1|^2 + |\gamma_2|^2$, are found similar for all triples: there are no difference between triples, which involve either (i) or (c) categories. In contrast, when maximum interference is assumed, the difference between (i) and (c) categories becomes more evident. For triples $(I{,}AB(i){,}B(i))$, the means of $|\gamma_1|^2$ and of $|\gamma_1|^2 + |\gamma_2|^2$ drop visibly. For triples $(I{,}AB(c){,}B(c))$, the average of $|\gamma_1|^2$ remains essentially unchanged (it decreases by just 0.03 in absolute value), and the mean of $|\gamma_1|^2 + |\gamma_2|^2$, while moderately decreasing, remains distinctly higher than for the alternative (i) category.

To formally ascertain the above observations, two-sample t-tests for equal means, without assumption of equal variances, were conducted. The hypotheses are formulated as follows:

- H_0: the coefficients $|\gamma_1|^2$ (or $|\gamma_1|^2 + |\gamma_2|^2$) for triples $(I{,}AB(i){,}B(i))$ and $(I{,}AB(c){,}B(c))$ come from independent random samples from normal distributions with *equal means*;
- H_1: the coefficients $|\gamma_1|^2$ (or $|\gamma_1|^2 + |\gamma_2|^2$) for triples $(I{,}AB(i){,}B(i))$ and $(I{,}AB(c){,}B(c))$ come from distributions with *unequal means*;
- H_1^* (a modified one-side hypothesis): the mean of coefficients $|\gamma_1|^2$ (or $|\gamma_1|^2 + |\gamma_2|^2$) for $(I{,}AB(i){,}B(i))$ *is smaller* than the corresponding mean for $(I{,}AB(c){,}B(c))$.

Results of the t-tests for equal means are summarized in table 6. For $\omega = 1$, the previous observation is in general confirmed. Under minimum interference condition, H_0 of equal means of $|\gamma_1|^2$ (or

$|\gamma_1|^2 + |\gamma_2|^2$) for triples ($I$,$AB(i)$,$B(i)$) and ($I$,$AB(c)$,$B(c)$) cannot be rejected, with p-values as high as 0.907, and the lowest p-value being 0.127. At the same time, for the maximum interference case, the p-values of H_0 are 4 to 6 times smaller than for the minimum interference case and, for $|\gamma_1|^2 + |\gamma_2|^2$, the p-values of H_0 decrease to p-value=0.03, when pitting H_0 against H_1^*.

For $\omega = 0.5$, the tendency is similar for the means of $|\gamma_1|^2$: H_0 cannot be rejected, but the p-value for the maximum interference case is almost 4 times lower than for the minimum interference case, signaling an increasing discriminability between (i) and (c) types. The means of $|\gamma_1|^2 + |\gamma_2|^2$ under both conditions - minimum and maximum interference - can be considered unequal with reasonable significance levels: H_0 can be rejected with p-value=0.10 (resp., 0.05) in favor of H_1 (resp., H_1^*). However, the condition $\omega = 0.5$ should be analyzed with caution, as many of the raw calculated coefficients required adjustments, which may lead to some distorsions.

Table 7 provides results of a similar two-sample t-test, but now conducted not between two types of triplets - (I,$AB(i)$,$B(i)$) and (I,$AB(c)$,$B(c)$), - but rather within each type. The test investigates the susceptibility of triple's types - (i) and (c) - to a change of interference, from the minimum to the maximum interference case. Analyzing $|\gamma_1|^2$, for triples with compatible (c) categories, the hypothesis H_0 of equal means is convincingly not rejected while, for triples with incompatible (i) categories, H_0 can be rejected at a significant level (with maximum p-value= 0.02). Thus, triples (I,$AB(i)$,$B(i)$) are distinguished by their susceptibility to a change of interference.

Table 7: Results of two-sample t-tests for equal means (without assumption of equal variances) within each triple type, under changing interference conditions. H_0: means of $|\gamma_1|^2$ (or $|\gamma_1|^2 + |\gamma_2|^2$) within one type of triples - either (I,$AB(i)$,$B(i)$), or (I,$AB(c)$,$B(c)$) - are *equal*; H_1: means of $|\gamma_1|^2$ (or $|\gamma_1|^2 + |\gamma_2|^2$) within one type of triples - either (I,$AB(i)$,$B(i)$), or (I,$AB(c)$,$B(c)$), - are *unequal*; H_1^* (a modified one-side hypothesis, reported in brackets): mean of $|\gamma_1|^2$ (or $|\gamma_1|^2 + |\gamma_2|^2$) for the case of maximum interference *is smaller* than for the case of minimum interference. The table shows that triples (I,$AB(i)$,$B(i)$) can be distinguished by their susceptibility to a change of interference.

	Susceptibility of triple's types - (i) and (c) - to a change of interference:							
	$\|\gamma_1\|^2$				$\|\gamma_1\|^2 + \|\gamma_2\|^2$			
	(I,$AB(i)$,$B(i)$)		(I,$AB(c)$,$B(c)$)		(I,$AB(i)$,$B(i)$)		(I,$AB(c)$,$B(c)$)	
For $\omega = 1$:	μ	p-value	μ	p-value	μ	p-value	μ	p-value
Interference case:								
minimum (zero)	0.73	**0.048**	0.74	**0.757**	0.84	0.003	0.90	0.080
maximum	0.55	**(0.024)**	0.71	**(0.379)**	0.60	(0.002)	0.77	(0.040)
For $\omega = 0.5$:	μ	p-value	μ	p-value	μ	p-value	μ	p-value
Interference case:								
minimum (zero)	0.71	**0.089**	0.69	**0.892**	0.91	0.0006	0.97	0.002
maximum	0.55	**(0.044)**	0.68	**(0.446)**	0.65	(0.0003)	0.77	(0.001)

To summarize the main observation on aggregated coefficients, under two interference conditions - minimum (zero) and maximum - triples with incompatible (i) categories behave differently, than triples with compatible (c) categories. Vice versa, the comparison of the coefficients' dynamics under different interference cases may help to classify prospects, e.g. (i) versus (c) type.

In our above analyses of different conditions - minimum and maximum interference, $\omega \in \{1, 0.5\}$, - the same empirical data is used and the general analytical formulation of prospects and decision maker's state of mind are kept unchanged. Our analysis thus reveals how the experimental data *could* be explained under each condition, and whether and to what degree the influence of different factors (decision modes and interference between them) varies. In this context, a decrease of the weight $|\gamma_1|^2$ of the predominant (typicality) decision mode (and the subsequent decrease of the sum

$|\gamma_1|^2 + |\gamma_2|^2$), which is observed for the maximum interference case (relative to zero interference), is associated with the increased influence of the QDT interference factor q under this condition. Recall that the attraction factor q reflects the interference between decision modes of a prospect which, after exposure to the prospect, are also present in the state of mind. In addition, the q-factor may include interferences contributed by other elementary prospects of the state of mind, which are originated, not from the prospect in consideration but, from other sources. The latter corresponds to the general QDT formulation of a state of mind as a statistical operator, or would require the identification of additional elementary prospects within the current formulation in terms of a pure state of mind. However, whatever the source, interference effects are condensed into the QDT attraction q-factor.

The analysis of this section shows that the decrease of $|\gamma_1|^2$, followed by a growing effect of the q-factor, is mostly observed for the triples with incompatible (i) categories, and not for compatible (c) type. Thus, the observed probabilities (probability judgements) of prospects with (c) categories can often be explained with a relatively small q-factor. In contrast, in order to explain the probabilities of prospects with (i) categories, the q-factor plays a substantial role. Given the findings in section 5.1, triples $(I,AB(i),B(i))$ are characterized by a negative attraction factor q. Now an additional hypothesis can be formulated: with the decrease of the coefficient $|\gamma_1|^2$ of the predominant (typicality) decision mode, the magnitude of the negative attraction factor q is expected to increase. In other words, an increase $|\gamma_1|^2$ should lead to an increase of q.

To test this hypothesis, the adjusted coefficients $|\gamma_1|^2$ for the case of maximum interference are compared with the attraction factor q from table 2. Figure 6 illustrates this dependence. Values of $|\gamma_1|^2$ and q are analyzed for the case for $\omega = 1$. Firstly, this case is considered to be more reliable. It leads to fewer outliers among γ coefficients, see table 10 in appendix A.3. In addition, as was mentioned in section 5.1, the case of $\omega = 1$ seems to be more plausible, since it recovers closely the observed amplitudes of the conjunction fallacy and the conjunction effect, which were reported in table 1. Secondly, for the value of $\omega = 1$, the attraction factor q from table 2, i.e. from equation 60, reduces to a simple form

$$\begin{aligned} q(\pi_{(AB)_1 I}) &= p_{exp}(AB) - 0.5 \\ (resp., \quad q(\pi_{B_1 I}) &= p_{exp}(B) - 0.5\ ). \end{aligned} \tag{64}$$

Indeed, the attraction factor from equation (64) can be more generally understood as a refinement of a prospect probability (probability judgement) from a simple toss-like guess (50/50) by incorporating within the q-factor the additional information about a subject (I) and one of four categories ($AB(i)$, $B(i)$, $AB(c)$, $B(c)$).

Figure 6 confirms the hypothesis. As the predominant (typicality) decision mode is characterized by a larger coefficient $|\gamma_1|^2$ (≥ 0.5), i.e. the decision (probability judgement) is determined, to a larger degree, by a single major factor, the QDT attraction factor q varies in a wide range (-0.4, 0.3). However, when the influence of the predominant decision mode decreases and $|\gamma_1|^2 < 0.5$, i.e. the uncertainty of a decision increases, than the magnitude of the negative attraction q increases. The next section elaborates this observation.

As also vividly illustrated by figure 6, in the region of higher certainty (right side), triples with compatible categories $(I,AB(c),B(c))$ are characterized by a moderate positive attraction factor. The gradual increase of uncertainty flips the q-factor to the negative side, however the absolute magnitude of the interference effect remains relatively small. This confirms our aggregate analysis, and demonstrates the mechanism in a finer manner.

Triples with incompatible categories $(I,AB(i),B(i))$ has been shown (table 7) to exhibit a significant decrease of their $|\gamma_1|^2$ coefficients, as well as sums of the coefficients, when going from the maximum

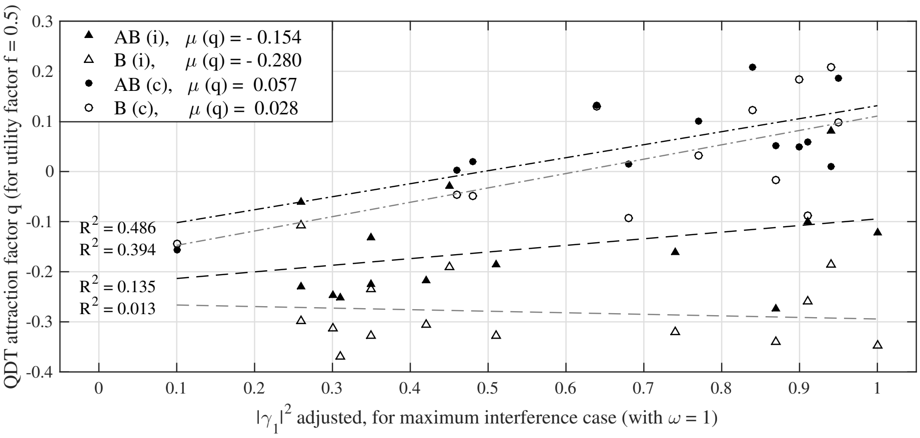

Figure 6: Dependence between the QDT attraction factor q and the average squared probability amplitudes $|\gamma_1|^2$ of a predominant (typicality) decision mode $|(AB)_1 I_1\rangle$ of a prospect. A prospect is a probability judgement of an instance I (a subject) to belong to one of the four categories: an incompatible conjunction $AB(i)$ (filled triangles), or its constituent $B(i)$ (empty triangles), a compatible conjunction $AB(c)$ (filled circles), or its constituent $B(c)$ (empty circles). The figure illustrates the case of maximum interference between typicality $|(AB)_1 I_1\rangle$ (resp. $|B_1 I_1\rangle$) and atypicality $|(AB)_2 I_1\rangle$ (resp., $|B_2 I_1\rangle$) decision modes. Values of q and $|\gamma_1|^2$ are presented for the case of $\omega = 1$. For each category type, sample mean of the attraction factor $\mu(q)$ is provided in the inset.

to the minimum interference condition, leading to growing negative interferences and q-factors. With increased uncertainty, the negative attraction factor of prospects with $AB(i)$ categories further decreases. However, within a conjunction, category A is typical of an instance I, and provides a compensating (positive) effect on the overall negative q-factor. This keeps the q-factor of a prospect within the conjunction category $AB(i)$ larger than for its constituent prospect with a single incompatible $B(i)$ category.

The negative q-factors of prospects within the incompatible constituent category $B(i)$ consistently exhibit the largest amplitude (average $q(B(i)) = -0.28$), of the order of and even exceeding the prediction of the QDT "quarter law" (see section 3.1). This raises interesting questions, to be explored in the future, regarding possible lower values of the q-factor, the existence of a lower barrier, and its insensitivity to a change of the uncertainty level.

This completes the second interpretation of conjunction fallacy.

6 From "attraction" q to universal (uncertainty, risk and loss) "aversion" q

A remarkable insight from figure 6 is that, as the uncertainty of a prospect increases, all linear regressions, although relatively noisy (maximum $R^2 = 0.49$), converge to the range of $q \in (-0.25, -0.15)$. This convergence is observed for all prospects, regardless of their type. A negative attraction factor for prospects with incompatible (i) categories could be expected. But the q-factor of prospects with compatible (c) categories also converges to the same *negative* range. We propose to refer to this universal convergence value of the attraction factor (equiv., convergence range (-0.25,-0.15)) as the "aversion" q.

The value of this universal "aversion" q resembles the QDT "quarter law" (see section 3.1), which predicts an average (among participants or repetitions) absolute value $|q| = 1/4$. The quarter law suggests an identical average value of q in both directions - positive and negative - and independently of uncertainty (or of the average uncertainty level). In contrast, the observed "aversion" q is the result of a general tendency observed for *any type of prospect* to converge to the same *negative* range at high *high uncertainty* levels and independently of the (un-)attractiveness of a prospect under more certain conditions.

The universal "aversion" q provides a theoretical and empirical background for the use of a general QDT-based "uncertainty aversion principle". According to Yukalov and Sornette (2009), this principle suggested that the uncertain prospect alternative is associated with the most negative attraction factor. This was shown to lead to reasonable agreements with top-level aggregated empirical results. However, no justification of such an assumption was provided, and the choice of a prospect with higher uncertainty was quite arbitrary. The current empirical analysis reveals a much more complicated and subtle picture than previously assumed. The universal convergence of the QDT attraction factor to $q \in (-0.25, -0.15)$ that we have documented here is shown to be associated with increased uncertainty, and thus resembles closely the well-known *uncertainty aversion.* It actually can be understood as an interpretation and mechanism of uncertainty aversion within the QDT framework, which in addition provides quantitative predictions, e.g. within the convergence range $q \in (-0.25, -0.15)$.

Going one step further, different types of aversions - uncertainty, risk, loss aversion, - which have been thoroughly discussed in "classical" decision theories, may have the same core repulsion mechanism that manifests itself (slightly differently) under different conditions (more/less uncertainty; gain/loss domain). Scattered evidence of the possibility to quantify different risk attitudes with the QDT attraction factor q were provided in previous studies:

- evidence of larger risk aversion of females compared to males (Favre et al., 2016),
- confirmation of a gender risk aversion effect, and evidence of a "safe q" that is based on standard deviation risk measure (Siffert et al., 2017),
- integration of loss aversion with "large loss" q (Vincent et al., 2017).

For the first time, the current study may be able to provide a common theoretical basis for modelling different risk attitudes with the QDT attraction factor. The various well-documented risk preferences may originate from the same interference mechanism and be modelled with universal principles: an irreducible "QDT indeterminacy (uncertainty) principle", and the universal "aversion" q .

7 Conclusions

In this article, we have taken a fresh look at a classical example of behavioral patterns, the conjunction fallacy, which turned out to be insightful. Our reinterpretation of the conjunction fallacy within Quantum decision theory (QDT) clarified the distinction between two origins of interfering elementary prospects (decision modes):

- different prospects associated with a choice situation (which is at the core of previous QDT developments);
- the state of mind that characterizes the decision maker (background, knowledge, experience, psychological traits, feeling, emotions, and so on).

This distinction does not contradict, but rather aims at building on top and complementing previous QDT developments. We clarify that interfering decision modes may be present (incepted, framed) in the *state of mind*, prior (and during) to the occurrence of a choice situation, and those interfering elementary prospects may *affect* the representation of the choice prospects (options). Thus, not only choice options influence the constituents (elementary prospects) of a state of mind, as previously considered, but the reverse dependence should also be taking into account. The second source of interference can be useful to understand many of the observed behavioral patterns and cognitive biases, such as: representativeness and availability heuristics, which were previously invoked to explain the conjunction fallacy (Tversky and Kahneman, 1983), as well as anchoring, attentional bias, belief bias, confirmation bias, framing effect, optimism and pessimism biases, etc. The different cognitive biases should thus been classified according to the two sources of interference.

In this article, the conjunction fallacy was reinterpreted and quantitatively analysed from two perspectives involving either

- the uncertainty factor (uncertainty angles) - figure 2; or
- the probability amplitudes of interfering decision modes - figure 6.

For the first time, an in-depth quantitative analysis within QDT was performed. Previous studies involved top-level aggregate experimental results, exploiting the most general QDT relation: $p = f + q$. They either checked agreement of data with that relation, or sought for parametrization of f and q, based on “classical” decision theories. The current study aimed at exploring the interference mechanism of QDT, based on the quantification of two fundamental elements: a decision maker’s state of mind and the prospects under consideration. The quantification of the decision maker’s state of mind was achieved by introducing a link between typicality judgements and probability amplitudes of decision modes in the state of mind. Dividing the problem into several (extreme) cases - minimum and maximum interference, as well as varying the weights of interfering modes in the state of mind - allowed us to estimate the uncertainty and relative contributions of prospect’s decision modes to their probability judgement. This level of granularity enabled the analysis of broader and more detained datasets, where interference effects are less profound (for example, data on compatible prospects). This also opens the possibility of discovering the interdependencies and dynamics of QDT elements, and of making the theory operational.

Our explanation of the conjunction fallacy remains squarely based on the core idea of QDT, that uncertainty is the source and modulating factor of the interference (q) between decision modes, which affects the probability of a prospect to be chosen (i.e. a probability judgement). We showed that the interference factor q is present for all types of prospects, conjunctions (both compatible and incompatible) and their constituents. However, it manifests itself differently. For example, prospects with compatible (c) categories are found less susceptible to a change of interference conditions (minimum or maximum interference), and the coefficients of interfering decision modes vary within a smaller range. For this type of prospects, the q-factor is more likely to be characterized by a smaller absolute amplitude, which may switch between positive and negative values depending on the level of uncertainty (i.e. relative contribution of the predominant decision mode). For prospects with incompatible (i) categories, interference effects are more easily detected, because they are more likely to be negative for any uncertainty level. Thus, they are usually characterized by a larger absolute q-factor. Within an incompatible conjunction $AB(i)$, we found that a category A, which is formulated such that it is typical of an instance I, then provides a compensating (positive) effect on the overall negative q-factor of the prospect. In contrast, for a prospect with an incompatible constituent category $B(i)$, the negative q-factor has the largest amplitude, with an average $q(B(i)) = -0.28$, which is at the order and even exceeds the prediction of the QDT “quarter law” in previous studies. Issues concerning the lowest possible values of the q-factor, the possible existence of a lower "barrier", and the insensitivity of such prospects to a change of the uncertainty level, are important future research directions.

Based on our detailed empirical analysis, the novel principle of a "QDT indeterminacy (uncertainty)" has been introduced, as being the fundamental limit to the precision with which certain pairs (sets) of prospects can be simultaneously known (assessed) by a decision maker, or elicited by an experimental procedure.

The observed general tendency, for *any type of prospect*, of the QDT attraction factor to converge to the same *negative* range $q \in (-0.25, -0.15)$ for *high uncertainty* levels motivated the introduction of an universal "aversion" q. The "aversion" q was found essentially independent of the (un-)attractiveness of a prospect under more certain conditions. This is contrast with the QDT "quarter law" previously introduced (Yukalov and Sornette, 2009, 2016b), We have provided supporting evidence for the *QDT uncertainty aversion principle* and clarified its application.

For the first time, the present study may be able to provide a theoretical common ground for modelling different risk attitudes - uncertainty-, risk-, loss-aversion, - with the QDT attraction factor. These various behavioral characteristics are well documented by "classical" decision theories in different choice conditions. We suggest that they may originate from the same interference mechanism and explained with universal principles: the irreducible "QDT indeterminacy (uncertainty) principle" and the universal "aversion" q.

A Appendices

A.1 Experimental results from (Shafir et al., 1990).

Table 8: Each of the 14 instances I (description of a subject) was presented to decision makers on separate occasions with one of 4 categories: either incompatible (i) conjunction and its constituent, or compatible (c) conjunction and its constituent. 110 participants made judgements about typicality (typ_{exp}) of an instance I in each category (54 participants), and probability (p_{exp}) that an instance I belonged to the corresponding category (distinct 56 participants). Difference between typicality (resp., probability) judgements with respect to a conjunction (AB) and its constituent (B) are referred to as conjunction effect (resp., conjunction fallacy)

Instance I	Type	Conjunctive category AB:		Constituent category B:		Conjunction effect:	Conjunction fallacy:
		$typ_{exp}(AB)$	$p_{exp}(AB)$	$typ_{exp}(B)$	$p_{exp}(B)$	$typ_{exp}(AB)-typ_{exp}(B)$	$p_{exp}(AB)-p_{exp}(B)$
1	i	0.439	0.401	0.264	0.241	0.175	0.160
	c	0.733	0.601	0.64	0.533	0.093	0.068
2	i	0.340	0.274	0.125	0.173	0.215	0.101
	c	0.773	0.707	0.664	0.623	0.109	0.084
3	i	0.252	0.226	0.177	0.160	0.075	0.066
	c	0.445	0.344	0.294	0.355	0.151	-0.011
4	i	0.337	0.367	0.133	0.266	0.204	0.101
	c	0.716	0.686	0.623	0.598	0.093	0.088
5	i	0.289	0.269	0.121	0.202	0.168	0.067
	c	0.729	0.552	0.686	0.484	0.043	0.068
6	i	0.249	0.282	0.110	0.194	0.139	0.088
	c	0.604	0.550	0.675	0.683	-0.071	-0.133
7	i	0.389	0.377	0.161	0.152	0.228	0.225
	c	0.772	0.578	0.799	0.544	-0.027	0.034
8	i	0.383	0.252	0.208	0.188	0.175	0.064
	c	0.564	0.516	0.400	0.407	0.164	0.109
9	i	0.527	0.471	0.195	0.310	0.332	0.161
	c	0.694	0.559	0.538	0.412	0.156	0.147
10	i	0.375	0.314	0.144	0.172	0.231	0.142
	c	0.671	0.684	0.668	0.642	0.003	0.042
11	i	0.559	0.580	0.282	0.315	0.277	0.265
	c	0.750	0.519	0.540	0.452	0.210	0.067
12	i	0.340	0.249	0.078	0.131	0.262	0.118
	c	0.575	0.509	0.779	0.707	-0.204	-0.198
13	i	0.291	0.339	0.116	0.180	0.175	0.159
	c	0.668	0.503	0.512	0.453	0.156	0.050
14	i	0.423	0.439	0.229	0.392	0.194	0.047
	c	0.679	0.633	0.670	0.629	0.009	0.004

A.2 Verification of the necessary conditions for the emergence of the QDT attraction factor

The necessary conditions for the attraction factor to be non-zero are: (a) entanglement in a strategic decision-maker state, and (b) entanglement of a prospect, i.e. a decision is to be made under uncertainty. Concerning the former condition (a), a strategic decision-maker state can be separable (not entangled) only if the measurements of observables are not temporally correlated. Subsection 4.3.1 demonstrates the evolution of a strategic decision-maker state through a sequence of channels, which produces an entangled state $\hat{\rho}_{ABIM}$ (resp., $\hat{\rho}_{BIM}$) just prior to a decision. For the second condition (b), to determine whether prospect $\pi_{(AB)_1 I}$ from (28) is entangled, we investigate the separability of the corresponding operator $\hat{P}_{(AB)_1 I}$ (30), as proposed in (Yukalov and Sornette, 2015, 2016a). For this, we introduce two Hilbert-Schmidt spaces below. The first one is defined by

$$\widetilde{\mathcal{AB}} \equiv \{\mathcal{AB}, \mathcal{H}_{AB}, \sigma_{AB}\} \tag{65}$$

where $\mathcal{AB} = \{\hat{P}_{(AB)_i}\}$ is an operator algebra (or an algebra of local observables), acting on the Hilbert space $\mathcal{H}_{AB}$, while σ_{AB} is the scalar product $\sigma_{AB} : \mathcal{AB} \times \mathcal{AB} \to \mathbb{C}$ that is defined as

$$\sigma_{AB} : (\hat{P}_{(AB)_1}, \hat{P}_{(AB)_2}) = \mathrm{Tr}_{AB} \hat{P}^{\dagger}_{(AB)_1} \hat{P}_{(AB)_2} \tag{66}$$

and generates the Hilbert-Schmidt norm $\|\hat{P}_{(AB)}\| \equiv \sqrt{(\hat{P}_{(AB)_i}, \hat{P}_{(AB)_i})}$, $i \in 1,2$. Similarly, for the second Hilbert-Schmidt space:

$$\widetilde{\mathcal{I}} \equiv \{\mathcal{I}, \mathcal{H}_I, \sigma_I\} \tag{67}$$

where $\mathcal{I} = \{\hat{P}_{I_i}\}$ is an operator algebra (or an algebra of local observables) on the Hilbert space $\mathcal{H}_I$, the scalar product is a map $\sigma_I : \mathcal{I} \times \mathcal{I} \to \mathbb{C}$ that is defined as

$$\sigma_I : (\hat{P}_{I_1}, \hat{P}_{I_2}) = \mathrm{Tr}_I \hat{P}^{\dagger}_{I_1} \hat{P}_{I_2} \tag{68}$$

and generates the Hilbert-Schmidt norm $\|\hat{P}_I\| \equiv \sqrt{(\hat{P}_{I_i}, \hat{P}_{I_i})}$, $i \in 1,2$. Now, a composite Hilbert-Schmidt space can be introduced as a tensor-product space

$$\widetilde{\mathcal{AB}} \otimes \widetilde{\mathcal{I}} = \{\mathcal{AB}, \mathcal{H}_{AB}, \sigma_{AB}\} \otimes \{\mathcal{I}, \mathcal{H}_I, \sigma_I\}. \tag{69}$$

An operator acting on this composite Hilbert-Schmidt space $\widetilde{\mathcal{AB}} \otimes \widetilde{\mathcal{I}}$ is called *separable* (or disentangled) if and only if it can be represented as a sum

$$\sum_i \hat{P}_{(AB)_i} \otimes \hat{P}_{I_i} \qquad (\hat{P}_{(AB)_i} \in \widetilde{\mathcal{AB}}, \quad \hat{P}_{I_i} \in \widetilde{\mathcal{I}}) \tag{70}$$

Importantly, the operator $\hat{P}_{(AB)_1 I}$ cannot be represented in the separable form (70), because the last term $|I_k\rangle \langle I_l|$ in (30) does not pertain to $\widetilde{\mathcal{I}}$. Thus, we conclude that the corresponding composite prospect $\pi_{(AB)_1 I}$ in (28) is *entangled*.
Following the same procedure, the operator $\hat{P}_{B_1 I}$ in (31) is non-separable, i.e. it cannot be represented as a sum

$$\sum_i \hat{P}_{B_i} \otimes \hat{P}_{I_i} \qquad (\hat{P}_{B_i} \in \widetilde{\mathcal{B}}, \quad \hat{P}_{I_i} \in \widetilde{\mathcal{I}}) \tag{71}$$

and the related composite prospect $\pi_{B_1 I}$ in (28) is *entangled* as well.

Thus, the necessary conditions for the emergence of the QDT attraction factor in (36), i.e. $q(\pi_{(AB)_1 I}) \neq 0$ and $q(\pi_{B_1 I}) \neq 0$, are satisfied.

A.3 Squared probability amplitudes $|\gamma_i|^2$, $i \in \{1,2\}$, of prospects' decision modes

Table 9: Adjusted coefficients $|\gamma_i|^2, i \in \{1,2\}$, and their sums, for the Minimum interference case. The coefficients are calculated for all 28 triples of an instance I and categories AB and B, including 14 incompatible (i) and 14 compatible (c). Adjustments are: exclusion of outliers (marked by a dash); and corrections to satisfy the constraint $\Re_{\geq 0} \in [0,1]$, i.e. replacement of negative values by 0, and renormalization of $|\gamma_1|^2$ and $|\gamma_2|^2$, if their sum exceeds 1. Identified by the general ESD test, outlier values are marked by an asterisk *. Unadjusted values are in brackets.

Minimum (zero) interference, $\cos\Delta(\pi) = 0$							
I	Type (AB,B)	$\omega = 1$			$\omega = 0.5$		
		$\lvert\gamma_1\rvert^2$	$\lvert\gamma_2\rvert^2$	$\sum_{i=1,2} \lvert\gamma_i\rvert^2$	$\lvert\gamma_1\rvert^2$	$\lvert\gamma_2\rvert^2$	$\sum_{i=1,2} \lvert\gamma_i\rvert^2$
1	i	0.91	0.00	0.91	0.91	0.00	0.91
	c	0.80	0.07	0.86	0.80	0.13	0.93
2	i	0.58	0.11	0.70	0.58	0.23	0.81
	c	0.88	0.11	0.99	0.80 (0.88)	0.20 (0.22)	1.00 (1.10)
3	i	0.88	0.00	0.89	0.88	0.01	0.89
	c	0.30	0.38	0.68	0.29 (0.30)	0.71 (0.75)	1.00 (1.06)
4	i	0.70	0.20	0.90	0.63 (0.70)	0.37 (0.40)	1.00 (1.10)
	c	0.95	0.01	0.96	0.95	0.02	0.97
5	i	0.55	0.15	0.71	0.55	0.31	0.86
	c	– (0.98)	– (-0.60)*	– (0.38)	– (0.98)	– (-1.20)*	– (-0.22)
6	i	0.76	0.12	0.88	0.75 (0.76)	0.25	1.00 (1.01)
	c	– (1.29)	– (-0.58)*	– (0.71)	– (1.29)	– (-1.16)*	– (0.13)
7	i	0.98	0.00 (-0.01)	0.98 (0.97)	0.98	0.00 (-0.01)	0.98 (0.97)
	c	– (0.29)	– (1.55)*	– (1.84)*	– (0.29)	– (3.10)*	– (3.39)*
8	i	0.48	0.11	0.59	0.48	0.22	0.70
	c	0.81	0.14	0.95	0.74 (0.81)	0.26 (0.28)	1.00 (1.09)
9	i	0.70	0.22	0.92	0.62 (0.70)	0.38 (0.43)	1.00 (1.13)
	c	0.85	0.00 (-0.09)	0.85 (0.75)	0.85	0.00 (-0.19)	0.85 (0.66)
10	i	0.70	0.08	0.78	0.70	0.17	0.87
	c	– (5.29)*	– (-8.71)*	– (-3.42)*	– (5.29)*	– (-17.42)*	– (-12.13)*
11	i	0.96 (1.00)	0.04 (0.05)	1.00 (1.05)	0.92 (1.00)	0.08 (0.09)	1.00 (1.09)
	c	0.60	0.28	0.88	0.52 (0.60)	0.48 (0.56)	1.00 (1.16)
12	i	0.55	0.10	0.64	0.55	0.19	0.74
	c	0.92	0.00 (-0.05)	0.92 (0.87)	0.92	0.00 (-0.10)	0.92 (0.82)
13	i	0.93 (0.98)	0.07	1.00 (1.06)	0.87 (0.98)	0.13 (0.15)	1.00 (1.13)
	c	0.61	0.29	0.90	0.51 (0.61)	0.49 (0.58)	1.00 (1.19)
14	i	0.58	0.34	0.92	0.46 (0.58)	0.54 (0.67)	1.00 (1.25)
	c	0.70 (0.78)	0.30 (0.33)	1.00 (1.11)	0.54 (0.78)	0.46 (0.66)	1.00 (1.44)
μ	i	0.73 (0.74)	0.11	0.84 (0.85)	0.71 (0.74)	0.21 (0.22)	0.91 (0.96)
	c	0.74 (1.10)	0.16 (-0.49)	0.90 (0.60)	0.69 (1.10)	0.28 (-0.98)	0.97 (0.11)
	all	0.74 (0.92)	0.13 (-0.19)	0.87 (0.73)	0.70 (0.92)	0.23 (-0.38)	0.93 (0.54)
σ	i	0.17 (0.18)	0.09	0.14 (0.15)	0.18	0.16 (0.19)	0.10 (0.16)
	c	0.20 (1.24)	0.14 (2.42)	0.09 (1.20)	0.22 (1.24)	0.25 (4.84)	0.05 (3.61)
	all	0.18 (0.89)	0.12 (1.71)	0.12 (0.85)	0.19 (0.89)	0.20 (3.41)	0.09 (2.55)
IQR	all	0.30 (0.34)	0.17 (0.20)	0.12 (0.24)	0.33 (0.34)	0.30 (0.40)	0.11 (0.29)

Table 10: Adjusted coefficients $|\gamma_i|^2, i \in \{1,2\}$, and their sums for the maximum interference case. The coefficients are calculated for all 28 triples of an instance I and categories AB and B, including 14 incompatible (i) and 14 compatible (c). Adjustments are: exclusion of outliers (marked by dash); and corrections to satisfy the constraint $\Re_{\geq 0} \in [0,1]$, i.e. replacement of negative values by 0, and renormalization of $|\gamma_1|^2$ and $|\gamma_2|^2$, if their sum exceeds 1. Identified by a general ESD test, outlier values are marked by an asterisk *. Unadjusted values are in brackets.

Maximum interference, $\cos\Delta(\pi) = 1$							
I	Type (AB,B)	$\omega = 1$			$\omega = 0.5$		
		$\|\gamma_1\|^2$	$\|\gamma_2\|^2 2$	$\sum_{i=1,2} \|\gamma_i\|^2$	$\|\gamma_1\|^2$	$\|\gamma_2\|^2$	$\sum_{i=1,2} \|\gamma_i\|^2$
1	i	0.91	0.00	0.91	0.91	0.00	0.91
	c	0.77	0.00	0.77	0.77	0.00	0.77
2	i	0.35	0.05	0.39	0.35	0.10	0.44
	c	0.84	0.01	0.85	0.84	0.01	0.85
3	i	0.87	0.00	0.87	0.87	0.00	0.87
	c	0.10	0.26	0.36	0.10	0.51	0.61
4	i	0.35	0.10	0.45	0.35	0.21	0.56
	c	0.95	0.00	0.95	0.95	0.00	0.95
5	i	0.26	0.08	0.34	0.26	0.17	0.43
	c	0.87 (1.31)	0.13 (0.20)	1.00 (1.52)	– (1.31)	– (0.41)	– (1.72)*
6	i	0.42	0.06	0.48	0.42	0.11	0.54
	c	0.90 (1.73)	0.10 (0.20)	1.00 (1.93)	– (1.73)	– (0.40)	– (2.13)*
7	i	1.00	0.00	1.00	1.00	0.00	1.00
	c	– (0.12)	– (0.92)*	– (1.04)	– (0.12)	– (1.84)*	– (1.96)*
8	i	0.30	0.04	0.35	0.30	0.08	0.39
	c	0.68	0.02	0.70	0.68	0.05	0.72
9	i	0.45	0.08	0.53	0.45	0.17	0.62
	c	0.91	0.01	0.92	0.91	0.01	0.92
10	i	0.51	0.02	0.54	0.51	0.05	0.56
	c	– (28.26)*	– (37.83)*	– (66.09)*	– (28.26)*	– (75.66)*	– (103.92)*
11	i	0.94	0.00	0.94	0.94	0.01	0.95
	c	0.48	0.06	0.54	0.48	0.12	0.60
12	i	0.31	0.05	0.36	0.31	0.09	0.40
	c	0.94	0.00	0.95	0.94	0.00	0.95
13	i	0.74	0.02	0.76	0.74	0.04	0.78
	c	0.46	0.07	0.53	0.46	0.14	0.60
14	i	0.26	0.19	0.45	0.26	0.38	0.64
	c	0.64	0.06	0.70	0.64	0.11	0.76
μ	i	0.55	0.05	0.60	0.55	0.10	0.65
	c	0.71 (2.73)	0.06 (2.83)	0.77 (5.56)	0.68 (2.73)	0.10 (5.66)	0.77 (8.39)
	all	0.62 (1.64)	0.05 (1.44)	0.68 (3.08)	0.60 (1.64)	0.10 (2.88)	0.70 (4.52)
σ	i	0.28	0.05	0.24	0.28	0.11	0.21
	c	0.26 (7.36)	0.08 (10.08)	0.21 (17.43)	0.27 (7.36)	0.16 (20.15)	0.14 (27.50)
	all	0.28 (5.23)	0.06 (7.13)	0.24 (12.35)	0.28 (5.23)	0.13 (14.27)	0.19 (19.49)
IQR	all	0.52 (0.57)	0.08 (0.09)	0.46 (0.47)	0.53 (0.57)	0.12 (0.17)	0.32 (0.36)

Table 11: Results of the generalized (extreme Studentized deviate) ESD test [a] to detect $i = 1...k$ outliers in samples of γ coefficients, approximated by the normal distribution. The test statistic $R_i = (max_i|x_i - \overline{x}|)/s$, where $\overline{x}$ is the sample mean and s is the sample standard deviation, is provided. The critical value λ_i is calculated based on the t-distribution. Two-sided test was performed with significance level $\alpha^{GESD} = 0.001$, for each coefficient $|\gamma_i|^2$, $i \in \{1, 2\}$, and their sum, and under each condition (minimum and maximum interference, and $\omega \in \{1, 0.5\}$). The values of the detected outliers are presented, and the corresponding triples (I, AB, B) are listed in I_{out} by an index number of an instance I and a type of categories AB, B - compatible (c) or incompatible (i).

Minimum (zero) interference, $\cos\Delta(\pi) = 0$														
$\|\gamma_1\|^2, \omega \in \{1, 0.5\}$			$\|\gamma_2\|^2, \omega = 1$			$\|\gamma_2\|^2, \omega = 0.5$			$\sum_{i=1,2} \|\gamma_i\|^2, \omega = 1$			$\sum_{i=1,2} \|\gamma_i\|^2, \omega = 0.5$		
R_i	λ_i	Outlier	R_i	λ_i	Outlier	R_i	λ_i	Outlier	R_i	λ_i	Outlier	R_i	λ_i	Outlier
4.939	3.464	5.29	4.990	3.464	-8.71	4.990	3.464	-17.42	4.885	3.464	-3.42	4.973	3.464	-12.13
2.360	3.441		3.924	3.441	1.55	3.927	3.441	3.10	3.838	3.441	1.84	4.108	3.441	3.39
2.189	3.416		2.918	3.416	-0.60	2.924	3.416	-1.20	2.844	3.416		3.359	3.416	
2.438	3.390		3.594	3.390	-0.58	3.605	3.390	-1.16	2.011	3.390		3.306	3.390	
I_{out}: 10(c)			I_{out}: 10(c), 7(c), 5(c), 6(c)			I_{out}: 10(c), 7(c), 5(c), 6(c)			I_{out}: 10(c), 7(c)			I_{out}: 10(c), 7(c)		

Maximum interference, $\cos\Delta(\pi) = 1$														
$\|\gamma_1\|^2, \omega \in \{1, 0.5\}$			$\|\gamma_2\|^2, \omega = 1$			$\|\gamma_2\|^2, \omega = 0.5$			$\sum_{i=1,2} \|\gamma_i\|^2, \omega = 1$			$\sum_{i=1,2} \|\gamma_i\|^2, \omega = 0.5$		
R_i	λ_i	Outlier	R_i	λ_i	Outlier	R_i	λ_i	Outlier	R_i	λ_i	Outlier	R_i	λ_i	Outlier
5.090	3.464	28.26	5.101	3.464	37.83	5.101	3.464	75.66	5.100	3.464	66.09	5.101	3.464	103.92
2.855	3.441		4.589	3.441	0.92	4.589	3.441	1.84	3.214	3.441		2.939	3.441	2.13
2.212	3.416		2.719	3.416		2.646	3.416		2.846	3.416		3.229	3.416	1.96
1.674	3.390		2.366	3.390		2.410	3.390		1.557	3.390		3.511	3.390	1.72
I_{out}: 10(c)			I_{out}: 10(c), 7(c)			I_{out}: 10(c), 7(c)			I_{out}: 10(c)			I_{out}: 10(c), 6(c), 7(c), 5(c)		

---

[a] A generalized ESD test is done with Matlab *gesd* function, which was created by F.A. Alcaraz Garcia, and based on (Rosner, 1983).

Bibliography

D. Aerts, J. Aerts Arguëlles, Lester Beltran, Lyneth Beltran, M. Sassoli de Bianchi, S. Sozzo, and T. Veloz. Testing quantum models of conjunction fallacy on the world wide web. *International Journal of Theoretical Physics*, 2017.

J.R. Busemeyer, E.M. Pothos, R. Franco, and J.S. Trueblood. A quantum theoretical explanation for probability judgment errors. *Psychological review*, 118(2):193, 2011.

M. Favre, A. Wittwer, H.R. Heinimann, V.I. Yukalov, and D. Sornette. Quantum decision theory in simple risky choices. *PLoS ONE*, 12(11), 2016.

R. Hertwig and G. Gigerenzer. The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of behavioral decision making*, 12(4):275, 1999.

B. Rosner. Percentage points for a generalized esd many-outlier procedure. *Technometrics*, 25(2): 165–172, 1983.

E.B. Shafir, E.E. Smith, and D.N. Osherson. Typicality and reasoning fallacies. *Memory & Cognition*, 18(3):229–239, 1990.

M. Siffert, T. Kovalenko, and D. Sornette. Calibration of quantum decision theory, analysis of parameters, gender difference and risk aversion. Working paper, ETH Zurich, Zurich, 2017.

A. Tversky and D. Kahneman. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90(4):293, 1983.

S. Vincent, T. Kovalenko, V.I. Yukalov, and D. Sornette. Calibration of quantum decision theory, aversion to large losses and predictability of probabilistic choices. *Submitted to Theory and Decision*, 2017.

J.B.W. Webber. A bi-symmetric log transformation for wide-range data. *Measurement Science & Technology*, 2(24), 2013.

S.S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9:60–62, 1938.

V.I. Yukalov and D. Sornette. Processing information in quantum decision theory. *Entropy*, 11(4): 1073–1120, 2009.

V.I. Yukalov and D. Sornette. Quantum probabilities of composite events in quantum measurements with multimode states. *Laser Physics*, 23(10):105502, 2013.

V.I. Yukalov and D. Sornette. Manipulating decision making of typical agents. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(9):1155–1168, 2014.

V.I. Yukalov and D. Sornette. Positive operator-valued measures in quantum decision theory. In Harald Atmanspacher, Claudia Bergomi, Thomas Filk, and Kirsty Kitto, editors, *Quantum Interaction: 8th International Conference, QI 2014, Filzbach, Switzerland, June 30 - July 3, 2014. Revised Selected Papers*, pages 146–161. Springer International Publishing, Cham, 2015.

V.I. Yukalov and D. Sornette. Quantum probability and quantum decision-making. *Phil. Trans. R. Soc. A*, 374(2058):20150100, 2016a.

V.I. Yukalov and D. Sornette. Inconclusive quantum measurements and decisions under uncertainty. *Frontiers in Physics*, 4(12):(doi: 10.3389/fphy.2016.00012), pp.1–9, 2016b.

4 Conclusion

This Chapter specifies the main contributions, some of the blank and blind spots of the conducted research and delineates perspective on future directions.

As for the resilience strand of research, we attempted to embrace very different systems and scientific areas, to identify commonalities between dissimilar approaches to resilience and to systematize them in a general framework. As our main contributions we perceive:

- an interpretation of resilience as a, complementary to risk, measure of the stress of a system, which dynamically characterize its (potential) reaction to (adversary, exo/endo) stressors;
- a four-level resilience hierarchy, which represents an inclusive relation: engineering resilience $\subset$ ecological resilience $\subset$ viability $\subset$ adaptability/transformability;
- a framework - "4 quadrants" of risk severity and system control, - which identifies four risk and resilience regimes, the corresponding response mechanisms and management tools.

We hope that this general framework can be useful as a guidance and a top-level design instrument for a holistic Risk-Resilience (R-R) management system. It can help to foresee possible regimes of a system, to develop adequate measurement techniques, to adapt tactics and strategy, and to take timely countermeasures. For future developments, the theoretical framework allows for an (almost direct) mapping between the regimes of "4 quadrants" and the four levels of resilience hierarchy. Thus, an efficient resilience methodology should be put into correspondence and channeled towards each functional regime.

At the same time, we acknowledge limitations and challenges that the deployment of an ambitious R-R system would unavoidably face. Some of the common issues were outlined. However, despite providing numerous examples, references and focus on socio-economic systems, we did not undertake a detailed case study of a particular system. As we argued that there is no silver bullet to all types of stress-factors, we neither advocate for an one-for-all-sizes resilience solution, which would fit any system. Thus, a great effort is required to transform the general framework into an implementable R-R management system with tailored instruments, methods and processes. As a next step to tackle these practical issues, industry-specific standardization of resilience processes and their gradual/selective integration into an existing risk management system should be considered.

Regarding Quantum decision theory (QDT), we emphasize the following contributions:

- a delineation of the evidence of the intrinsically probabilistic decision making process: the investigation of a probabilistic model of choice reversal and intrinsic limits of choice predictability;
- the first QDT parametrization on a mid-size dataset of individual and aggregate binary risky choices, with separation of the aversion to large losses into an interference q-factor;
- a novel interpretation of the conjunction fallacy, which invokes a decision maker's state of mind as a distinct source of uncertainty and interference effects;

- the first in-depth data-driven quantitative analysis of weights of interfering decision modes within a state of mind and within a prospect, and their interconnections;
- propositions of an universal "aversion" factor $q \in (-0.25, -0.15)$ under high uncertainty levels, and the "QDT indeterminacy principle".

We are enthusiastic about these findings, which contribute to the theoretical development of QDT and bring the theory a step closer to being operationalized.

Several blank spots and limitations should be considered. The parametrization of QDT based on the two components of the prospect's probability - utility f-factor and attraction q-factor - can be continued. We have proposed only one of the possible combinations, incorporating stochastic cumulative prospect theory and the constant absolute risk aversion function, which attributes aversion to large losses to an interference effect q. Naturally, other "classical" decision theories, as well as new analytical factors can be tested. The advantage of this approach is in a direct relation of QDT to the traditional decision field, and in the possibility to separate different risk attitudes in the q-factor. However, the disadvantage of this approach is that it focuses only on the most general result of QDT, just scratching the surface of the theory. We see the realization of the full potential of QDT through the investigation of its underlying interference mechanism, which may allow for a deeper understanding of the decision making process and truly original insights.

Our inferences were based on empirical datasets from two experiments - (i) binary risky choices and (ii) typicality/probability judgements. The experimental setups were designed to be incentive compatible, to avoid undesired (order and representation) effects, and include two types of measurements - repeated, i.e. within group, in (i); and independent, i.e. between groups, in (ii). Nevertheless, behavioral experiments can hardly be absolutely free from criticism. QDT, maybe more than other theories, highlights the complexity of a decision maker's state of mind and the emergent property of a decision. Quantification and control of the factors that influence human choices is challenging, especially due to the feedbacks in the state of mind. The laboratory environment itself may serve not as an isolating condition, but as a strong interfering factor that effects decisions. Thus, conclusive results require a scrupulous analysis of numerous different experiments, as well as empirical results obtained from humans in the wild.

Keeping in mind these shortcomings, we encourage further exploration of the universal "aversion" q-factor and its connection to other risk attitudes (aversion to ambiguity, uncertainty, risk and loss), which should include various experimental setups. Elicitation of personal traits and noninvasive analysis of additional data sources (e.g. online and mobile activity, social media) are of particular interest to QDT. We encourage study of memory effects in repeated experiments with between-repetition intervals of different duration. Among interesting exploratory questions are - influence of learning, information exchange and interactions between individual decision makers. The formulation and testing of the "QDT indeterminacy principle" is another intriguing direction of research.

Our dream is to make QDT practically useful, for example by developing a decision support tool, which incorporates interference effects of multiple factors, or a quantum scoring application.

To conclude, we see resilience as an emergent property of a system, which relies on decision making. The social (human) factor plays a determining role in many systems of interest. Thus, decision making, at different levels, is an essential element of the resilience build-up. For an individual (and from the perspective of psychology and decision theory), interconnection and even inseparability of resilience and sound decision-making are self-evident. For a large multilayer system, its resilience (including social resilience) emerges bottom-up, from the strength of its individual elements. On the other hand, a top-down design and management of a resilient system is supported by decision-making tools. The importance of Resilience increases in an uncertain turbulent environment. Under the very same conditions, Quantum decision theory becomes distinctively relevant, providing instruments for uncertainty quantification and the natural account of fluctuating interfering factors.