U.PORTO

FEP FACULDADE DE ECONOMIA UNIVERSIDADE DO PORTO

Motivated Beliefs: Strategic Ignorance and Selective Memory in Investment Decisions

Hugo Teixeira França

Dissertation

Masters in Economics

Supervised by

João Correia da Silva

2018

Acknowledgements

I would like to present a special appreciation to Professor João Correia da Silva, for all the time spent orienting me, first on the discovery of the subject, and later on the construction of the dissertation. Thank you very much, because this thesis would not have been achieved without your guidance. I would also like to express my special thanks to the two tutors who guided me during these 2 years of masters: to Professor Vitor Carvalho for his great availability to help me find the one who seemed to be, and well, the best adviser to my thesis, and to Professor Joana Resende for all her relevant comments when the thesis was in the embryonic phase. The final thank-you goes to my family, my girlfriend and all my colleagues and friends who have offered me support and motivation.

Abstract

The need to give more realism to the way economic models represent human behavior is one of the main motivations for the development of this work. Here, I will address beliefs by their relevance to meeting psychological and functional needs, particularly as they shape individual motivation for investment decisions. This analysis is only possible, since the individuals present in this model are "psychologically enriched" by a set of human imperfections (imperfect self-knowledge, imperfect willpower and imperfect memory).

Thus, I will defend beliefs as a tool, not just a consequence of receiving information. Through the use of game theory, I will analyze how individuals engage in strategies for manipulating beliefs, namely through strategic ignorance and memory manipulation. It has been found that individuals often benefit from manipulating their beliefs, giving them motivation to execute projects in situations where they would not do so because of the dynamic inconsistency. Concretely, ex-ante manipulation of beliefs has a positive effect on individual motivation whenever individuals do not (a posteriori) intend to execute a project because of their dynamic inconsistency, but in fact, their ability is sufficient for them to do it. With respect to ex-post belief manipulation, it may turn against the agent. It is a strategy that can lead individuals to a situation where they never invest, even when they do not have negative news. Individuals depend on their memory, but if their credibility is zero, they have to resort to what they considered to be their ability if they had no information. However, since he would not invest in possessing what he considered to be his initial ability, individuals would rather not have the ability to manipulate his memory.

Key words: Multidisciplinary Analysis, Motivated Beliefs, Decision-Making, Rationality.

JEL codes: A12, D81, D83, D91.

Resumo

A necessidade de imprimir maior realismo à forma como os modelos económicos representam o comportamento humano é umas das principais motivações para o desenvolvimento deste trabalho. Aqui, tratarei as crenças pela sua relevância para satisfazer necessidades psicológicas e funcionais, nomeadamente na medida em que estas moldam a motivação individual para decisões de investimento. Esta análise só é possível, uma vez que os individuos presentes neste modelo são "psicologicamente enriquecidos" por um conjunto de imperfeições humanas (autoconhecimento imperfeito, força de vontade imperfeita e memória imperfeita).

Assim, defenderei as crenças como uma ferramenta, e não apenas uma consequência da recepção da informação. Através da utilização da teoria dos jogos, analisarei como os individuos incorrem em estratégias de manipulação das crenças, nomeadamente através da ignorância estratégica e da manipulação da memória. Verificou-se que os individuos beneficiam, frequentemente, da manipulação das suas crenças, conferindo-lhes motivação para executar projectos em situações em que não o fariam devido à inconsistência dinâmica. Concretizando, a manipulação ex-ante das crenças, têm um efeito positivo na motivação individual sempre que os individuos não pretendam (à posteriori) executar um projecto devido à sua inconsistência dinâmica, mas em que, na verdade, a sua habilidade é suficiente para o fazerem. Relativamente à manipulação das crenças ex-post, pode virar-se contra o agente. É uma estratégia que pode levar os indivíduos para uma situação em que este nunca investe, mesmo quando não tem nóticias negativas. Os individuos dependem da sua memória, mas se a credibilidade da mesma for nula, estes têm de recorrer àquilo que consideravam ser a sua habilidade caso não tivessem informação. Porém, uma vez que não investiria se possuisse aquilo que considerava ser a sua habilidade inicial, os individuos prefeririam não ter a capacidade de manipular a sua memória.

Palavras Chave: Análise Multidisciplinar, Crenças Motivadas, Tomada de decisão, Racionalidade.

Códigos JEL: A12, D81, D83, D91.

Table of Contents

1. Introduction 1

2. Literature Review 5

2.1. Behavioral Economics 5

2.1.1. An Overview 5

2.2. What are Motivated Beliefs and how they arise? 7

2.2.1. Beliefs 7

2.2.2. Motivated Beliefs 8

2.2.3. Psychologically evolved agents 9

2.3. Why are Motivated Beliefs important and pursued? 12

2.3.1. Why *Motivated Beliefs* are maintained? 12

2.3.2. How *Motivated Beliefs* are maintained? 14

2.4. Extensions 15

2.4.1. Internal Commitment Devices - Personal Rules 15

2.4.2. Social Influences on Self-Control 16

2.4.3. Intrinsic and Extrinsic motivation 17

2.5. Critiques to the game-theoretic approach of Motivated Beliefs 18

3. Motivated Beliefs Model 20

3.1. Distinctive Features 20

3.1.1. Self-Control and Dynamic Inconsistency 20

3.1.2. Rational decision-making paradigm 24

3.2. Ex-ante Belief Manipulation Strategies: Learn or Avoid? 25

3.2.1. Model Formal Features 25

3.2.2. Avoiding vs Learning Information 28

3.2.3. Four distinctive cases 29

3.2.4. Learn or avoid information? 31

3.3. Ex-post Belief Manipulation Strategies: Register or Forget? 34

3.3.1. Additional Formal Features 34

3.3.2. Not Registering vs Registering Negative News 37

3.3.3. Two Distinctive Cases 37

4. Conclusion 42

4.1. Future work 43

5. References 45

List of Figures

Figure 1 – Ex-ante Belief Manipulation Strategies 26

Figure 2 – Case 1: Self 1 always invests (no conflict of interest) 29

Figure 3 – Case 2: Self 1 does not invest if news are bad (conflict of interest) 30

Figure 4 – Case 3: Self 1 does not invest if news are bad (no conflict of interest) 30

Figure 5 – Case 4: Self 1 only invests if news are good (possible conflict of interest) 31

Figure 6 – Ex-post Belief Manipulation Strategies 36

Figure 7 – Ex-ante success probability 39

"A crença em Deus assenta em o que podemos chamar um acto de fé racional. Consciente ou inconscientemente, o movimento do espírito é este: (1) Tudo quanto existe é efeito de uma causa; o universo existe; portanto o universo é efeito de uma causa. (2) O efeito não pode conter mais que o que está contido na causa, (pois então seria efeito de mais causas que uma); o universo, no mais alto ponto em que nós o conhecemos, que é o homem, contém a consciência; portanto a causa do universo deve conter a consciência, isto é, deve ser uma Causa consciente. (3) O efeito não pode conter tudo quanto se contém na causa, pois então seria idêntico à causa, e não haveria causa nem efeito; o universo é múltiplo, extenso (no tempo e no espaço, ou no espaço-tempo) e diverso (isto é, composto de coisas não só muitas mas diferentes entre si); portanto a causa do universo tem que conter mais que multiplicidade, ou seja totalidade, mais que extensão, ou seja infinidade, mais que diversidade, ou seja plenitude. Cumpre advertir que totalidade se diferencia de plenitude em que o primeiro é um conceito quantitativo, o segundo qualitativo: assim a totalidade do prazer seria a soma de todos os prazeres possíveis, a plenitude do prazer a concentração em um só prazer do que se acha contido na diversidade de todos.

Por qualquer especulação desta ordem, em geral subconsciente ou instintiva, chega o homem à crença racional na existência de Deus. Que é racional, já o vimos, não esqueçamos porém que é simples crença, pois parte de princípios naturais, instintivos, mas dialecticamente contestáveis.

Organizado, como é, o espírito do homem, não há demonstrações senão a científica, isto é, a que se baseia ou na observação, ou na experimentação, ou no cálculo, ou em qualquer combinação destas três coisas. Ora, ainda admitindo que o conceito de causa e efeito seja induzível da observação (o que é contestável e, de facto, tem sido contestado), o que é certo é que o que chamamos universo em seu «conjunto» não é susceptível de observação, de experimentação ou de cálculo, pois não temos sentido algum com que o abranjamos, nem sabemos, portanto, o que em esse «conjunto» (e já conjunto é hipótese) o universo seja.

A existência de Deus é, pois, indemonstrável, mas é um acto de fé racional, natural portanto — inevitável até — em qualquer homem no uso da sua plena razão.

E tanto assim é que o ateísmo anda sempre ligado a duas qualidades mentais negativas — a incapacidade de pensamento abstracto e a deficiência de imaginação racional. Por isso, nunca houve grande filósofo ou grande poeta que fosse ateu.

Indemonstrabilidade dos dados dos sentidos — a crença que temos na realidade objectiva do universo é um acto de fé sensual. A ciência mostra a que leis obedecem os seres que povoam esse universo aparentemente objectivo; não demonstra, nem pode demonstrar que essa objectividade seja mais que aparente, que o mundo não seja sonho e ilusão. Num caso há um acto de fé logicamente injustificável, na validade da razão; no outro um acto de fé, igualmente injustificável, na validade dos sentidos. É bom que o ateu saiba que, se o teísta pratica uma imprudência lógica ao ocultar a existência de Deus, o ateu a pratica igual ao ocultar a existência de uma pedra, entendendo eu, por este segundo termo, a imaginação que sabe figurar-se «entes» despidos de todos os atributos."

1915

***Textos Filosóficos**. Vol. II Fernando Pessoa (Estabelecidos e prefaciados por António de Pina Coelho) Lisboa, Ática*, 1968, p.76

1. Introduction

In the economic sphere, extremely simplified models are often used, mainly based on simplistic behaviors and preferences. Economists are aware of these shortcomings that lead to the questioning, by other social scientists, of the hypotheses used (Tirole, 2002). However, what is lost in realism is a gain in predictive capacity, a feature extremely relevant to the role of economics among policy makers. Still, the simplifications seem to be too restrictive in a number of cases, which leads us to ask what benefits would be achieved by introducing a small piece of realism, provided it is introduced in a way that does not compromise the capabilities held by models with more simplistic assumptions (Tirole, 2002).

It is this need to introduce realism into the models that mimic human behavior that fundamentally motivates this incursion into the field of Behavioral Economics to work on Motivated Beliefs. Being the analysis of human behavior something that belongs to the realm of Psychology, this enters into economic reality given its importance for economic agents' decision-making. In the wide range of different issues that Behavioral Economics encompasses, we intend to develop a work that allows the mathematical operationality, always so present in economics, to help us draw important conclusions about the way in which agents motivate themselves by using rational strategies of belief manipulation (Bénabou & Tirole, 2002a).

In this thesis, the focus will be on beliefs for their contribution to completing psychologically and functionally the set of individual needs. (Bénabou & Tirole, 2016). I aim to answer a fundamental question and two complementary ones: Is it possible for a rational agent according to Bayes to manipulate and defend his beliefs? Through what mechanisms? And, it will always be better for this individual to manipulate his beliefs?

Here is fundamental to understand how beliefs, incorrect or not, are maintained, knowing in advance that "*the strength and resiliency of certain beliefs cry out for explanation* (Gilovich, 1991, p.2)". Consider what an example presented in the literature focused on human judgment and reasoning states: when a nurse believes that more babies are born on a full moon night, even if such belief is factually incorrect (Abell & Greenspan, 1979), it cannot be

said that it is for lack of exposure to the relevant evidence. As Gilovich stated "e*rroneous beliefs plague both experienced professionals and less informed laypeople alike* (Gilovich, 1991, p.2)".

Individuals do not hold beliefs out of stupidity, or because they are too gullible. In fact, the human being is even quite gifted with very precise intellectual tools in the processing of information, and many of the beliefs that are questionable arise from the "*misapplication or overutilization of generally valid and effective strategies for knowing* (Gilovich, 1991, p.2)". Thus, what are considered doubtful beliefs may be due exclusively to imperfections in our abilities, and thus have only cognitive origin[1]. This analysis may seem to reduce the reliability of individual beliefs. However, it is true that here fundamentally wrong beliefs are treated, but we cannot disregard the number of times individuals are right in their beliefs.

Thereby, at the turn of the millennium, developments in game theory and cognitive psychology made the distance from the rationality hypothesis partly counteracted by the emergence of certain forms of adaptability or at least of implicit purposefulness in human cognition (Bénabou & Tirole, 2016). This is what gives the starting point for the approach that will be presented subsequently. Thus, the work to be carried out will be developed on the basis of a *Motivated Beliefs* model built by Jean Tirole and Roland Bénabou[2], in which rationality is a fundamental assumption. It will be argued that decisions which sometimes seem to happen as a result of the lack of rationality of an agent are, in fact, motivated by agents who want to defend beliefs that are cherished to them, and so occur without moving away from their rationality (Bénabou & Tirole, 2002a).

Since we intend to evaluate the importance of motivated beliefs in an economic context, I will focus on this thesis in its importance for decision-making. Consequently, this thesis main objectives are to identify and explore intrapersonal conflict cases through the use of a game-theoretic approach, explaining how belief manipulation strategies help individuals to defend their self-confidence (important to motivate them to undertake projects). Here, this self-confidence will be assessed in the individuals' beliefs about their individual ability.

[1] Gilovich also argues that beliefs "*are the products, not of irrationality, but of flawed rationality* (Gilovich, 1991, p.3)".

[2] Based on Bénabou and Tirole (2002) and Bénabou (2013).

Once established the research question and the objectives that this thesis intends to identify are presented, I hope that the most valuable contributions that can come from this analysis are easily perceptible. Firstly, the integration of existing literature on Motivated Beliefs, focusing mainly on intrapersonal conflict caused by dynamic inconsistency, in a unified work that only explores this topic. Next, a formal analysis of how an individual manipulates his beliefs (ex-ante and ex-post the arrival of information) through his distinctive characteristics, relative to the standard rational paradigm. And, finally, understanding under what conditions, aiming at the defense of beliefs, individuals incur an intrapersonal conflict of interest since the optimal decision may depend on the moment of time.

In the model that will be presented here, I will try to show that in individual investment decisions, beliefs have a predominant influence. Here, first, I will look at how information-learning decisions influence the investment decisions taken, namely how they vary both according to the initial belief in individual ability. Evaluating in detail the most interesting cases, it was verified that with different initial beliefs different information structures are required for an individual to execute an investment.

In a second phase of this thesis one discusses, not the escape to the information, but the capacity to manipulate it after it has already been received. This manipulation does not have to be done directly, "eliminating" negative information, it can be accomplished by controlling attention and awareness at the moment information is received. This power to manipulate information will prove to be harmful to the individual. The initial belief in this case is defined at the outset, making the investment decision dependent solely on the manipulation of the subsequent belief (through manipulation of the info received). What will happen is that the ability to manipulate memory will lead an individual to a situation in which he never invests, this will happen because the credibility in his memory is zero. Since without this ability of manipulation an agent invested at least when he had no negative news, this ability is detrimental.

One of the challenges of a multidisciplinary approach is to "please" the different views that experts have in different areas. Logically, it will never be possible to tailor an approach in order to frame all contributions, so that only the best possible analysis within the chosen conceptual framework can be objectifiable.

In this thesis, in Chapter 2 will have the framing and definition of the theme of Motivated Beliefs, helping to perceive the relevance of the same, to understand how they arise and how they are pursued by individuals. At the end of this chapter some distinct strands of the Motivated Beliefs model are presented but will not be explored in depth here as well as a reflection on the main critiques of this approach. Then, in Chapter 3, we begin with a formal analysis of the relevant cases, after a brief exposition of the distinctive characteristics of the model, according to the moment when the initial information and beliefs of the individuals about their ability are received. Chapter 4 concludes and identifies possible directions for future work.

2. Literature Review

This section presents a state-of-the-art analysis of Motivated Beliefs in the context of Behavioral Economics, highlighting the role of issues such as self-control, beliefs and motivation. It begins in section 2.1, by framing the theme in the context of Behavioral Economics, exploring later in section 2.2, what are the beliefs, motivated beliefs, what characteristics individuals should be equipped to form them (and defend them) and how they do it. In section 2.3, we explain the importance of beliefs for individuals (for their motivation and satisfaction of needs), as well as the behaviors that lead to the formation and defense of beliefs. Then, will be explained some other extensions, in section 2.4, that this model can incorporate such as the connection between motivated beliefs and internal commitment devices (Personal Rules), the analysis to Motivated Beliefs group or social perspective and the way extrinsic and intrinsic motives can complicate individual motivation. Finally, I conclude, in section 2.5, with a brief account of the main criticisms to this type of approach, which may arise from the various "factions" of behavioral economists or from out of the profession.

2.1. Behavioral Economics

2.1.1. An Overview

This thesis lies within the scope of Behavioral Economics, which is a field that arises mainly from the approach to Psychology, and contrary to what is sometimes said, it is not a change in the methodology of economics, but a return to the way it began (Thaler, 2016). The field has grown incredibly in the last decades, but it did not arise in this period. Going back to the beginnings of economics, in particular to Adam Smith's *Theory of Moral Sentiments* of 1759, we find the starting point for a multidisciplinary analysis. Nevertheless, after the attention denoted by the classics, it has fallen into disuse after the marginalist revolution.

The return came in the 1950s when Herbert Simon presented the term Bounded Rationality, that in the author's words intended "*(…)to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man (…)* (Simon, 1955, p.99)*"*. Therefore,

in his works he reiterated the need to link economics with psychology: "*one is tempted to turn to the literature of psychology (...)* (Simon, 1955, p.99)".

From there, Behavioral Economics grew by two distinct "branches" that Ainslie (2016) called "*the two legs of Behavioral Economics* (Ainslie, 2016, p. 261)". One is the motivational (behavioral) side that discovered anomalies in individuals' choice, focusing on the analysis of dynamic consistency (or inconsistency) of preferences and motives, with the discussion about hyperbolic, quasi hyperbolic and exponential discount, and their influence on self-control problems. The other is the cognitive side focused on individuals' anomalies as a result of framing, this is the analysis of systematic errors (heuristics and biases) from which emerges prospect theory (Kahneman & Tversky, 1979)[3].

Although the growth of the field has taken place at the turn of the millennium, it remains an area that is flourishing. One of the reasons why Behavioral Economics is more than a fad is that it has grown in very different directions within these "branches", incorporating both enthusiasts and skeptics, with works in which the more skeptic have introduced few insights of psychology, as well as works done by the most enthusiastic and supportive of agents' irrationality. To better understand what is being done, some have compiled the small advances that have taken place in this branches (DellaVigna, 2009; Rabin, 1998, 2002).

This thesis is closer to the approach of those who have carefully worked the approach to psychology (Tirole, 2002), and just as they did I intend to represent individuals behavior in the best possible way, but contrary to the aforementioned perspective of Herbert Simon, nothing here will be bounded (human characteristics as: rationality, willpower, self-interest). Instead, these characteristics may be imperfect. In what I will analyze here, the mistakes agents commit, happen even within their rationality.

I will largely follow the works developed by Bénabou and Tirole, insofar as its broad approach is clear and perceptible in explaining the relationship between rationality and motivated formation of beliefs, and for whom distorted beliefs are the consequence of a cognitive bias. However, their work was not the precursor in the analysis of the distortion of

[3] In the first "branch" appeared works as: challenge of the assumption that agents' decisions are always optimal (Kahneman & Tversky, 1979). Within the second "branch" appeared works like: The existence of systematic (purely mechanical) errors of inference (Tversky & Kahneman, 1974) or renewed attention to the existence of other behaviors inconsistent with standard economic theory (Thaler, 1980).

beliefs, but they produced the first works that formed a unified theory about how agents form their beliefs in a motivated way (ex-ante and ex-post information arrival). Long before, other works such as Greenwald (1980) and Kruger & Dunning (1999) dealt with wrong beliefs, attributing egocentrism, incompetence or self-serving bias as their causes. More recently, studies have emerged that addressed the issues of distorted beliefs in an alternative way. Köszegi (2006) attributed to the taste of individuals for a positive self-image the reason for overconfidence, and authors as Benoît and Dubra (2011) used an alternative approach to assert that overconfidence is mostly apparent.

Beyond this, as this thesis predominantly focuses on the relationship between belief distortion and individual's self-control problem (dynamic inconsistency), I will explore Bénabou and Tirole's approach to it[4]. To describe what is meant by self-control, I will consider it the same way that Bénabou and Tirole did, as "*people's common tendency to succumb to short-run impulses at the expense of their long-run interests* (Bénabou & Tirole, 2004, p.849)" and that *"is generally seen as reflecting conflicting internal preferences, such as when the individual's current self overweighs the present relative to the future. (…) or, in the words of Adam Smith, self-command, defined as the deliberate and reasoned overriding of powerful impulses and appetites* (Bénabou & Tirole, 2004, p.849)". This approach also allows me to treat belief manipulation, an ex-ante and an ex-post analysis, as the capacity to avoid information they may receive or to manipulate their awareness in the face of the information they inevitably receive, respectively, having as their ultimate goal their motivation.

2.2. What are Motivated Beliefs and how they arise?

2.2.1. Beliefs

Before defining what is meant by Motivated Beliefs it is worth explaining what beliefs are, more broadly, and in the context of economics particularly. Beliefs are all that an individual believes, and whether they are correct or erroneous, the value that individuals attach to them is the primary reason for their being learned and defended. In economics, the

[4] This topic is also referred in economic literature as intertemporal bargaining (Ainslie, 2016), self-command, self-control or time inconsistency.

use of the term beliefs describes the probability associated with the occurrence of certain event.

2.2.2. Motivated Beliefs

It is often considered, in economics, that individuals make good use of available information in order to achieve correct views of reality, enabling them to achieve the best rewards (payoffs). However, research about individual beliefs revealed, over the last decades, a set of trends (bias) that would not be expected from a standard rational viewpoint (Sharot & Garrett, 2016). Examples of this are superiority illusion[5] (as drivers believe they conduct better and safer than average drivers (Svenson, 1981)), or unrealistic optimism[6] (as students believe they are less likely to benefit from negative events and more likely to benefit from positive events than their colleagues (Weinstein, 1980)). Experiments demonstrate that individuals protect their beliefs by rationalizing the evidence they receive. It is especially interesting to look at the fact that individuals fundamentally defend their beliefs when they are "*value-laden*" (Bénabou, 2015, p. 669), that is, when these beliefs have associated a positive or negative value. An example of this is political beliefs (Kahan 2013). It was found that the opinions expressed about gun control or climate change, that individuals assigned to certain parties demonstrate, are not influenced by the cognitive ability of individuals. Thus, being closer to a particular party does not imply that an individual has greater cognitive capacity for information processing. However, the more cognitively sophisticated individuals were the ones who most demonstrated more ideology-driven opinions (Kahan, 2013).

Thus, it is easy to understand that the value placed by individuals on their beliefs leads to the emergence of non-Bayesian behavior as a way of maintaining or reinforcing them (Eil & Rao, 2011; Korn et al. 2014), being this is a permanent process of assessing costs, benefits, and risks of maintaining positive self-views (or world views) given the continued appearance of threatening evidence. It is these mental processes, which psychologists call motivated reasoning, that justify the use of the term Motivated Beliefs (Bénabou & Tirole, 2016).

---

[5] "*The tendency of the majority of people to believe they are better and more skilled than the majority of people* (Sharot & Garrett, 2016, p.26)".

[6] "*Overestimating the likelihood of future positive events and/or underestimating the likelihood of future negative events* (Sharot & Garrett, 2016, p.26)".

2.2.3. Psychologically evolved agents

One of the main contributions of the empirical experiments of Behavioral Economics to this work of Motivated Beliefs is the recognition that *Homo Economicus* is not always a good approximation of how individuals behave. Although it is not intended here to introduce all the imperfections that characterize *Homo sapiens*, it is intended to give realism to the way individuals, who populate economic models, behave. This is only possible by psychologically enriching these individuals, mainly with the addition of imperfections that display personal traits irrelevant so far. Hence, to allow beliefs defense, will be added a set of imperfections such as: *imperfect self-knowledge, imperfect willpower and imperfect recall* (Bénabou & Tirole, 2002a).

Firstly, it is very often that an individual derives "*what kind of a person*" (Bénabou & Tirole, 2011, p.807) he is through the choices he makes, since he struggles to know better his "Self" by observing those same choices (behaviors) (Bem, 1972). This brings us to *Imperfect Self-Knowledge,* as there is uncertainty about the true scope of individual ability, meaning that individuals do not know costs and benefits of their actions. This makes the decision of effort dependent on their self-confidence in their skills[7]. This feature counterbalances the emphasis that the standard rational paradigm places on the importance of signaling individual characteristics, since now the individual also have doubts about his "Self". Consequently, and according to psychology, a fundamental role of information seems to be that the individual knows himself, notwithstanding, individuals often seek to have the best possible self-view (Bénabou & Tirole, 2011). Introspective phenomena through which individuals struggle to understand their identity (self-esteem or self-justification) are examples of this (Bénabou & Tirole, 2002b). To understand how individuals, seek to use imperfect self-knowledge in a motivated way, it is worth analyzing an experience by Mazar, Amir and Ariely (2008). It was found that individuals like to consider themselves honest individuals. However, although they like to possess a good self-view[8], individuals understand that sometimes being dishonest pays off. So, what these researchers have discovered is that individuals "*are dishonestly enough to profit but honestly enough to delude themselves of their own integrity* (Mazar, Amir, & Ariely, 2008, p. 633)".

[7] It will not be surprising that they are not certain about the action to be taken, until the very moment.

[8] To deepen the understanding of self-views and self-esteem (Swann, Chang-Schneider, & Larsen McClarty, 2007).

Next, will be approached the *Imperfect Willpower*. Contrary to the standard rational economic individual, the individual described in this thesis is in a constant struggle to avoid succumbing to internal conflicts. For example, when an agent engages in a relationship or project (whatever its nature) it is critical to understand whether an individual can persevere in the face of adversity (Bénabou & Tirole, 2002b). Thus, this features often makes him unable to avoid impulses, sometimes leading to self-deceptions (namely, self-destructive behaviors). This directs us to the problem of dynamic inconsistency, and that in the model presented in this thesis will be represented through quasi hyperbolic discounting. In order to elucidate the reader of the set of situations in which this characteristic is verified in the individual behavior, it is worth to look at an example is set forth by Thaler, in 1981. In a laboratory experiment, this author found that median individuals were indifferent between $15 today and $20 a month thereafter (which resulted in a 345% annual discount rate) and between $15 today and $100 over 10 years (where the annual discount rate dropped to 19%). This suggests that because of the preference for immediate gratification, the discount is more pronounced in the near future than in the distant future (Thaler, 1981). Keeping in mind this preference reversal feature, it is easy to understand the role that past behavior monitoring has for individual self-control. However, due to the imperfection that follows, even though the problem of inconsistency is frequent, which could help an agent already know how to behave, he is not able to correctly remember his preferences.

Finally, one arrives at the last imperfection introduced in this model: *Imperfect Recall* (or *Motivated Cognition*). This feature allows the forgetfulness or distortion of a fraction of the memory[9], first by the limitations or decay and still by the existence of discretion in the information that will occupy the memory (conscious awareness). Firstly, contrary to standard rational models, reality has shown that there are several limitations that lead to the impossibility of having perfect memory, unlimited attention and full consciousness. In fact, agents suffer from a "hot-cold" empathy gap (Loewenstein, 1996) in such a way that they are unable to correctly recall through introspection ("cold" states) the intensity of short-term

[9] However, there are differentiated degrees of memory given that information can be more or less manipulable. In order of decreasing ease of manipulation we have: Past internal states (Loewenstein, 1996; Thaler et al., 1997), external circumstances and one's past actions (Bénabou & Tirole, 2002b).

feelings or temptations ("hot" states)[10], and this explains why agents are not able to accurately answer questions about the intensity of pain or temptation. Analyzing a relevant example, consider the case in which "*smokers who were not having a craving for a cigarette underestimated how much they would value a cigarette when they were later in a high craving state* (Sayette, Loewenstein, Griffin, & Black, 2008, p. 628)", what is verified is that once an individual is not experiencing the temptation he will almost certainly underestimate the influence of visceral factors (Loewenstein, 1996). Secondly, the fact that their memory is not perfect opens the door to a set of behaviors that evidence that individuals control their memory in a motivated way. It has been found that individuals have diverse ways of collecting performances or past actions in a self-serving manner: "*they tend to remember (be consciously aware of) their successes more than their failures, reframe their actions so as to see themselves as instrumental for good but not bad outcomes, and find ways of absolving themselves by attributing responsibility to others* (Bénabou & Tirole, 2002b, p. 140)". In a study of how individuals incorporate information (signals) that has intrinsic value (good or bad news) about their IQ, evidence of asymmetric updating has been found. It was found that in updating their beliefs, in response to the received signals, individuals who received good news updated their beliefs almost Bayesian, and that, in response to bad news, individuals ignored or strongly devalued the same information. This leads to beliefs very different from those that would be obtained with an update of beliefs according to a Bayes rule (Eil & Rao, 2011).

In literature the discretion issue seems to be the predominant one (Sharot & Garrett, 2016), once even when the information is perfectly reachable in memory the "*desirable and undesirable information is used differently to alter self-relevant beliefs, resulting in systematic biases*" (Sharot & Garrett, 2016, p. 25). So that, even without differences regarding the "freshness" of good or bad news in memory, it is evident the motivated cognition. However, there are also mechanisms or decisions that can be taken to avoid forgetting (or not) information, which helps to strengthen the existence of motivated cognition and opens the way for individuals to elaborate mechanisms for revisiting or not revising the information later (Bénabou & Tirole, 2002b).

[10] Or alternatively, "*when people are in an affective cold state (e.g., not experiencing thirst or craving at the moment), they do not accurately estimate how much such visceral states (hot states) will change their preference and behavior* (Sayette, Loewenstein, Griffin, & Black, 2008, p. 628)".

As a result of this enrichment of the model, it is possible to identify several personal traits such as: self-confidence, autonomy or dependency, willpower, self-reputation or intrinsic motivation. These traits will fill the set of reasons for individuals wanting to hold distorted beliefs (demand side of Motivated Beliefs). At the same time, this set of human imperfections also allows for the emergence of cognitive processes that explain how individuals pursue the goal of belief distortion (the supply side of Motivated Beliefs). These cognitive processes range from self-monitoring, personal rules, strategic ignorance or selective memory (Bénabou & Tirole, 2002b).

2.3. Why are Motivated Beliefs important and pursued?

The literature on beliefs is peremptory to say that individual beliefs fulfill an important set of psychological and functional needs (Bénabou & Tirole, 2016). The main emphasis in this thesis will be on the functional role of beliefs, namely on their importance to individual motivation (through increased self-confidence) given that individuals have imperfect willpower. Although some beliefs may be simply implausible or incorrect given the existing knowledge, many phenomena of motivated beliefs are very frequent (eg, overconfidence) and strongly influence individual decisions (Bénabou, 2015). These are powerful motivators, which help individuals overcome problems of dynamic inconsistency. In an investigation conducted by Puri and Robinson (2007) it has been found that the most optimistic individuals save more, retire later and are more likely to remarry. Thus, it is easy to understand that greater self-confidence helps, for example, that individuals do not procrastinate in the decision to save today or tomorrow. Taking into account what was previously stated, the two questions presented below represent the demand side (why) and supply side (how) of this Motivated Beliefs Model, respectively (Bénabou, 2015).

2.3.1. Why *Motivated Beliefs* are maintained?

This question, "*Why are Motivated Beliefs maintained?*", points us to the set of values that beliefs about individual's traits (self-confidence, self-esteem, autonomy or independence, etc.) have and which are the reasons why individuals intend to hold distorted beliefs (Sharot & Garrett, 2016). The above-mentioned values that beliefs possess include: a *consumption value*, a *signaling value* and a *motivation value*.

Individuals have a *consumption value* which comes from reflective consciousness of individuals, and which includes cases where they are happier simply by thinking well of themselves "as *when you lie awake in bed at night or look in the mirror glorying in your triumphs* (Tirole, 2002, p. 636)*"* or on the contrary, "*pride themselves on being ruthless businessmen, ultrarational economists, irresistible seducers, etc* (Bénabou & Tirole, 2002a, p. 906)". These are the cases where agents want to have a good self-image for purely affective issues which formally can be represented by a component of the utility function (being endogenously derived from "*fundamental preferences* (Bénabou & Tirole, 2002a, p. 906)"). Regarding the difficulty that could result from the amount of arguments that could be considered in the utility function, the solution is to consider "general ability" as the characteristic that encompasses all the possible hedonic arguments that individuals can identify as good for themselves. However, other difficulties emerge as there is no certainty regarding the appropriate shape of the hedonic preferences function[11]. This last obstacle to the formalization of the consumption value derived from agents' beliefs has been overcome with the introduction of a component (in individual's utility) that represents the anticipatory utility experienced when they think about their future prospects (however, just feelings felt before uncertainty resolution (Caplin & Leahy, 2001)), and it is built on a linear specification that dispenses the knowledge of the attitude towards the information (Bénabou, 2013; Koszegi, 2010). Nevertheless, this it will not be drawn in this thesis.

Information may also have a *signaling value*, this means that if an individual believes that his or her ability is high, it becomes easier to get others to believe. As explicitly stated by Bénabou and Tirole, "*it is often said that to lie most convincingly a person must believe his own lies. While the idea that people are "transparent" and have trouble misrepresenting their private information may seem unusual in economics, one could easily obtain an instrumental value of self-confidence from a signaling game where those who truly believe in their own abilities face lower costs of representing themselves favorably to others* (Bénabou & Tirole, 2002a, p. 877)".

Finally, there is *motivation value* (instrumental value of self-efficacy), which will be the most important value for the model elaborated in this thesis. This is a pivotal factor for the demand

[11] It could be growing and concave, which would imply that agents were information averse and that they had a positive self-esteem demand, or the opposite, in which agents would be information-lovers and the function would be decreasing and convex.

of self-confidence, which in turn is determinant to reach long-term goals. This value is what "*drags you out of bed, makes you undertake things and gets you going* (Tirole, 2002, p. 636)", that is, it is primarily responsible for fostering motivation in such a way that individuals persevere in their projects. In fact, confidence in one's abilities is an important factor "*as key to winning a medal, performing on stage, getting into college, writing a great book, doing innovative research, setting up a firm, losing weight, finding a mate, and so forth* (Bénabou & Tirole, 2002a, p. 877)". An instrumental analysis of self-confidence importance, as it will be done in this thesis, also evaluates the benefits of knowing individuals' limits and the costs of maintaining positive illusions.

2.3.2. How *Motivated Beliefs* are maintained?

Focusing now on the question "*how motivated beliefs are maintained?*", it is necessary to explore the set of cognitive processes through which individuals form and maintain distorted beliefs, that make up the supply side of self-esteem. It is, fundamentally, through two types of cognitive processes that belief distortion is pursued: information avoidance or information manipulation (affecting the probability of remembering information). Specifically, these cognitive processes include strategies like: strategic or blissful ignorance, and self-deception (reality denial)[12], respectively.

Firstly, strategic ignorance is an ex-ante belief manipulation strategy that lead to belief distortion. Here, are included different individual attitudes regarding information (e.g., ignorance). These strategies seek to counteract agents' orientation to the present. Since self-confidence is an individual "asset", there are cases where agents will not prefer to risk exposing themselves to new information, so that they do not fall into situations of procrastination because of their dynamic inconsistency (Carrillo & Mariotti, 2000). The formal presentation of this case will be made in Section 3.2.1.

Secondly, an individual can pursue belief manipulation through self-deception (or motivated cognition), an ex-post form of belief distortion. It is widely known, and multidisciplinarily documented, that there is a "*people's universal tendency to rehearse good news and to deny, explain away and selectively forget self-threatening information* (Bénabou & Tirole, 2002b, p. 145)". Thus, the most important issue is the analysis of how individuals behave (especially their memory) relatively to bad news, that will inevitably arise, about their ability or

[12] See Brown, Dutton, & Cook (2001) for a discussion of risks and benefits of self-deception.

performance (Bénabou & Tirole, 2002a). Within this concept of self-deception fit in several mechanisms such as: repression (avoiding individuals or situations that remember bad news), defensive denial (working hard not to think about problems) or self-serving attributions (such as "*linger over praise or positive feedback, rehearse them, and choose to be more frequently in environments or with people who will remind him of his past successes* (Bénabou & Tirole, 2002b, p. 146)"). However, it must be said that self-deception does not mean that agents can choose to possess the beliefs they want and this is because they cannot simply mechanically suppress memories. It happens that what controls agents' memory is their attention at the time of information arrival, which later provides selective memory or wishful thinking. This approach to self-deception follows those works of Bénabou and Tirole in which individuals, in order to be self-deceived, have to know and not know the same information simultaneously. What this means is that they have to maintain two opposing beliefs and, at the same time, not being aware of having one of them (Bénabou & Tirole, 2002b). Formal presentation of this examples will be made in Section 3.2.2.

2.4. Extensions

2.4.1. Internal Commitment Devices - Personal Rules

This thesis deals with the cognitive strategies that individuals use, and their relationship with some external commitment devices[13] that contribute to manipulate the perception of future individual reincarnations, through the elimination of threatening options. However, the literature on motivated beliefs is also capable of framing the relationship between cognitive strategies and internal forms of commitment to form "resolutions" or Personal Rules. Thus, the previously presented cognitive strategies (self-deception or strategic ignorance) that sought to converge short-term and long-term preferences, reducing time inconsistency, can also be used with internal commitment devices (or Personal Rules (Bénabou & Tirole, 2004)). These constitute a promise of future cooperation, and may exist on the form of: " *a diet, monthly savings targets; resolutions to smoke only after meals, jog twice a week, write five pages a day, always finish what one started, conduct one's life with dignity, and many similar "promises to oneself* (Bénabou & Tirole, 2004, p. 849)".

[13] Such as "*avoiding sources of temptation, holding illiquid assets, signing binding contracts and other forms of tying oneself to the mast*" or "*asking for controls* (Bénabou & Tirole, 2004, p. 849)".

According to Bénabou and Tirole, this relation between cognitive processes and forms of internal commitment described above represents a "*simultaneous choice of behavioral and information-processing principles* (Bénabou & Tirole, 2004, p. 876)". Additionally, resolutions are associated with credibility a key factor to individuals self-reputation. This happens fundamentally due to the human imperfections that this model incorporates. Since individuals memory is imperfect and they do not exactly know their willpower, their memory of past actions will define "*what kind of a person* (Bénabou & Tirole, 2004, p. 850)" an individual is. As individuals can only infer from their past actions, once motives have already been erased in the memory, sometimes even for cases that could be considered "exceptions" individuals prefer to remain faithful to their personal rule. There are a number of examples that can be highlighted, for example: when an individual continues to work even when he is very ill (or when a friend needs help), when the weather is terribly bad and an individual continues to go exercise or even when a friend offers you a dessert on a special occasion and you reject it, because of his diet (Tirole, 2002). In addition, in more worrying cases, individuals face "*workaholism, anorexia, miserness or oversaving (old people keeping the same saving rate even though they have no heirs)* (Tirole, 2002, p. 651)".

This set of examples referred above is important to understand how individuals can establish personal rules that avoid underregulation and at the same time end up falling into overregulation. These behaviors can lead individuals to self-traps. Agents are in a situation where they perceive their self-control depends on how they interpret present behavior and its influence on their future behavior. This implies that they are strongly averse to breaking rules because a lapse (opening of a precedent to impulsive behaviors) can create a "snowball effect" on their behavior, due to the loss of self-reputation (Bénabou & Tirole, 2004).

2.4.2. Social Influences on Self-Control

While it is intuitive to begin with the analysis of individual behavior explored in this thesis, the truth is that social and group phenomena are just as important to analyze (Bénabou & Tirole, 2002b). Individuals are naturally immersed in social relationships, which are an important factor influencing their beliefs. These relationships can lead to good or bad outcomes, especially when individual self-confidence (self-esteem) may be more fragile (Battaglini, Bénabou, & Tirole, 2005).

When questions of self-control become social phenomena, and individuals depend on collective or group thinking, the results of the appearance of distorted beliefs can be catastrophic. Bénabou (2013) explores the ways in which distorted beliefs emerge in the context of markets or companies and presents an interesting example: "*In the aftermath of corporate and public-sector disasters, it often emerges that participants fell prey to a collective form of willful blindness and overconfidence: mounting warning signals were systematically cast aside or met with denial, evidence avoided or selectively reinterpreted, dissenters shunned. Market bubbles and manias exhibit the same pattern of investors acting "color-blind in a sea of red flags* (Bénabou, 2013, p. 429)".

Also, when agents get into groups, they end up having a self-view useful to motivate them and that is capable to boost their performance (through manipulation). The main difference is that in this new social reality, manipulation is done by external elements (and not by previous "selves") and can be done in two different directions: the individual can try to influence others (self-presentation) or the social environment can to influence the individual ("looking-glass self") (Bénabou & Tirole, 2002b).

2.4.3. Intrinsic and Extrinsic motivation

Finally, there is still room to talk about another important issue in motivated beliefs approach. Since the extension to the social and group approaches allows individuals to be influenced by external motivators, it is worth taking some insights of the work, of 2003, developed by Bénabou and Tirole, on the subject. Here, questions such as the empowerment or rewards offered by a "principal" to agents and from which they draw important information for their decision-making can be evaluated. In this approach the relationship between individual's motivation and the way a principal (or the local environment) chooses the structure of the incentives that influence an individual is analyzed, "*from an economic and cognitive perspective* (Bénabou & Tirole, 2003, p.491)".

This approach analyzes the contexts in which the rewards are positive or negative reinforcers[14], and in which situations the intrinsic and extrinsic motivation can be conflicting in the formation and defense of beliefs. Following, in addition, clues of the existing literature: "*Indeed, a substantial body of experimental and field evidence indicates that extrinsic motivation (contingent*

[14] Sometimes, by the emergence of what the authors call "hidden costs" (Bénabou & Tirole, 2003).

rewards) can sometimes conflict with intrinsic motivation (the individual's desire to perform the task for its own sake) (Bénabou & Tirole, 2003, p. 490)".

Understanding these interactions between intrinsic and extrinsic motivation is important in several environments, as can be demonstrated by some examples as "*Gneezy and Rustichini (2000b) find that offering monetary incentives to subjects for answering questions taken from an IQ test strictly decreases their performance, unless the "piece rate" is raised to a high enough level. In the policy domain, Frey and Oberholzer-Gee (1997) surveyed citizens in Swiss cantons where the government was considering locating a nuclear waste repository; they found that the fraction supporting siting of the facility in their community fell by half when public compensation was offered* (Bénabou & Tirole, 2003, p. 490)". Thus, as it is easily understandable, it is important not to make one kind of incentives backfire the other one. Especially, to avoid that external incentives undermine individuals' self-confidence in the long run.

2.5. Critiques to the game-theoretic approach of Motivated Beliefs

This conceptual framework is obviously not immune to criticism. The main objective here is to express what the main shortcomings of Motivated Beliefs approach may be, in three distinct directions: focusing on those who are more skeptical of Behavioral Economics, on those who accept Behavioral Economics and who think that this model could go further, and on those who are adherents of the approach that highlights the irrationality of economic agents.

Firstly, lets focus on criticisms addressed by Behavioral Economics skeptics (also called "*rationalists* (Tirole, 2002, p. 641)"). These critics are not unaware, mostly, of standard rational economic paradigm failures, they are just skeptical of the path to where Economics can go as a result of the relationship with other disciplines. This happens because they fear that their discipline may lose its objectivity and parsimony. However, in this analysis it is not intended to enter into the vertigo of framing any and all cognitive biases (or human imperfections) that may possibly exist. Here, it will just be used a model that has a multidisciplinary concern, but without stop being pragmatic in the inclusion of psychology contributions. These should be added in such a way that improves the lessons that can be retrieved, as well as not resulting in the dilution of economic analysis traditional virtues.

Secondly, for those economists who may think that this conceptual framework is incomplete and that this approach could have gone further in the development of the topic, it is important to point out that the complexity of other extensions could lead to an excessive degree of difficulty that I am not sure would be useful for this economic analysis. One such case would be the inclusion of another human imperfection, namely the imperfect ability to assess probabilities. There is already an approach incorporating this imperfection, where it is included a formalization of it (Gabaix & Laibson 2005), which in any case is an important human imperfection since it prevents individuals from making really optimal decisions. Nevertheless, for the reason stated above, this imperfection will not be included in this work.

Finally, focusing on the critiques that may come from economists who emphasize evidences of agents' irrationality, I must mention that the introduction of these questions does not have room in this work. I intend to work on a model that only characterizes human behavior patterns, thus, this thesis does not seek to encompass such an individual view that draw conclusions for each individual idiosyncrasies (it is not addressed the cognitive shortcuts of our mind that "*operates automatically and quickly, with little or no effort and no sense of voluntary control* (Kahneman, 2011, p.16)", which Kahneman called "*System 1*(Kahneman, 2011, p.16)").

It is worth noting that the proposed approach of Motivated Beliefs is not the same one that deals with the transition from *Homo Economicus* to *Homo Sapi*ens as a result of the gradual loss of IQ (Thaler, 2000). Thus, it is intended to work on agent rationality (reflexivity and control provided by "*System 2* (Kahneman, 2011, p.16)" of our mind and that "*allocates attention to the effortful mental activities that demand it, including complex computations* (Kahneman, 2011, p.16)"). Even with the limitations I acknowledge with regard to this approach, I believe that this analysis is solidly defended and that deviations from the proposed approach, for this topic, can bring more difficulties than positive contributions to the analysis.

3. Motivated Beliefs Model

This model seeks to explain how belief manipulation happens in a motivated way, insofar as these beliefs arise within individual rationality. In this chapter I will firstly present distinctive model features which explain the need for individuals to maintain their motivation, through the defense of their beliefs. Then, the two forms of belief manipulation in which individuals may incur: strategic ignorance (section 3.2) or self-deception (section 3.3). This work is based on the contributions of Bénabou (2013, 2015), Bénabou and Tirole, (2002a), Carrillo and Mariotti (2000)[15].

3.1. Distinctive Features

3.1.1. Self-Control and Dynamic Inconsistency

As DellaVigna (2009) noted, individuals can depart from the standard rational paradigm on three issues: non-standard preferences, non-standard beliefs and non-standard decision making. The first class of departure from the standard rational theory can be found in this thesis, to the extent that this model of motivated beliefs includes the issue of dynamic inconsistency. The main reason for analyzing this problem is its importance in explaining the self-control problems that arise from individuals' dynamic inconsistency. Here, I'll focus on the discount factor between two distinct periods (whatever they are), since with dynamic inconsistency this discount factor depends on when the utility is evaluated as a result of individuals different preferences at different time periods.

The standard theory of intertemporal choice allows, through the use of an exponential (stationary) discount rate, that agents analyze and decide consistently intertemporally. With such a discount rate, it is considered that each individual always requires the same amount of reward for equal periods of time. Thus, any decision made today would be exactly the same if an individual would choose at any other period, since the transitivity of his choices

[15] This model will be based on these works, however some of the characteristics presented in these models will not be reproduced here.

would ensure that his preferences would remain unchanged (O 'Donoghue & Rabin, 1999)[16]. This consistency, a fundamental point for individual rationality, leads agents to optimum and consistent decisions at any point in time.

However, through the empirical observation of Economists, Psychologists, and Sociologists, it has been discovered and demonstrated that individuals are inconsistent in their dynamic decisions, thus calling into question the validity of the frequently used intertemporal choice theory. Strotz (1955) was one of the firsts to assert that the individuals presented inconsistencies intertemporally: "*An individual is imagined to choose a plan of consumption for a future period of time so as to maximize the utility of the plan as evaluated at the present moment. His choice is, of course, subject to a budget constraint. Our problem arises when we ask: If he is free to reconsider his plan at later dates, will he abide by it or disobey it - even though his original expectations of future desires and means of consumption are verified? Our answer is that the optimal plan of the present moment is generally one which will not be obeyed, or that the individual's future behaviour will be inconsistent with his optimal plan* (Strotz, 1955, p. 165)".

Multiple researches, subsequent to Strotz's work, have shown that individuals tend to prefer the present relative to the future, specifically due to the change in their implicit discount rate between periods. As Ainslie (1975) argued for the implicit discount rate, this "*has a pronounced upward concavity, suggesting that utility falls off rapidly for relatively short delays and declines more gradually as delays get longer* (Ainslie, 1975, p. 464)", and it is also verified that "*highly concave delay curves from some pairs of smaller-earlier and larger-later alternative rewards can cross, predicting an initial preference for the larger reward, which changes in favor of the smaller (specious) reward as the smaller reward becomes imminently available. This description accords with an intuitive view of what temptation is like* (Ainslie, 1975, p. 492)".

Experiments on animals and humans seem to prove Ainslie's statement. In experimentation with rats it was found that by offering them two alternative pairs of rewards, separated by certain intervals, the curves representing the implicit discount rates crossed. This was because, for the most part, one individual preferred one pellet immediately instead of three pellets 11 seconds later, but when the first pellet was only received 3 seconds after

[16] "*A person's relative preference for well-being at an earlier date over a later date is the same no matter when she is asked* (O 'Donoghue & Rabin, 1999, p.106)".

the initial moment, subjects preferred to wait 14 seconds to obtain the same 3 pellets (Logan, 1965). In another survey, with pigeons, and with a similar choice format, Chung and Herrnstein (1967) also obtained results in which individuals demonstrated a much higher implicit discount rate in the short term[17].

In the research on humans, Richard Thaler's work, "Some Empirical Evidence on Dynamic Inconsistency", provides evidence that "*subjects have a discount function which is non-exponential* (Thaler, 1981, p. 205)". In this work, a group of individuals was asked to state, in the hypothetical case of having won the lottery, the monetary value they would be willing to receive in the future to abdicate the reward in the present. What has been found, once again, is that individuals demanded a higher discount rate in the shorter waiting period[18]. Other authors evaluated the choices of undergraduates between shorter or longer periods of relief in the face of noxious noises (Solnick, Kannenberg, Eckerman, & Waller, 1980), and the tendency of pre-school-age children to choose between a meal that they want and another that they do not, depending on the waiting time to receive it (Mischel & Ebbesen, 1970), always having found a preference for the present. In DellaVigna's (2009) work it is also presented another set of good examples: "*When evaluating outcomes in the distant future, individuals are patient and make plans to exercise, stop smoking, and look for a better job. As the future gets near, the discounting gets steep, and the individuals engage in binge eating, light another (last) cigarette, and stay put on their job* (DellaVigna, 2009, p. 318)".

Thaler presents two possible explanations so that individuals do not present an exponential discount function between periods: on the one hand it may be a consequence of an individual arithmetical errors, or on the other hand, it may be because "*the psychophysics of time suggests that the difference between today and tomorrow will seem greater than the difference between a year from now and a year plus one day* (Thaler, 1981, p. 205)". Something that has also been found is that these inconsistencies arise both in naive individuals and in sophisticated individuals, although sophistication tends to mitigate the problem (O 'Donoghue & Rabin, 1999).

[17] To analyze other animal studies: Rachlin and Green (1972); Navarick and Fantino (1976); Ainslie and Herrnstein (1981). For a summary of several researches and experiments: Ainslie (1985).

[18] Also Shybut (1968) carried out investigations with money, in which they obtained results that corroborated those found by Thaler (1980).

This dynamic inconsistency problem is often addressed in economic literature as a problem of self-control. In fact, this is a highly valued issue in that dynamic consistency is the "motor" for individuals to control their impulses: to comply with diets, exercise plans or to be punctual (Laibson, 1997). As a way of ensuring their self-control, individuals often use forms of commitment (Strotz, 1955) in order to resolve possible intrapersonal conflicts arising from the reversal of individual preferences (retirement savings plans and deadlines to achieve certain goals are examples of this).

Exploring this problem, when individual preferences are dynamically inconsistent, it is common to construct a formal model in which individuals are represented by a sequence of temporal selves (reincarnations of the same individual at different periods of time) that make choices in a dynamic game. In order to model dynamic inconsistency (namely, "*bias for the present* (O 'Donoghue & Rabin, 1999, p. 106)"), it will be followed the approach developed by (Phelps & Pollak, 1968), which has since been used in many jobs (Dellavigna & Malmendier, 2004; DellaVigna & Pope, 2018; Laibson, 1997; O 'Donoghue & Rabin, 1999)[19]. As Bénabou and Tirole (2002a) said "*the resulting structure is that of a game of strategic communication between the individual's temporal selves* (Bénabou & Tirole, 2002a, p. 875)".

To model this problem as well as possible, I will follow Laibson's (1997) formalization. Will be used preferences of the type (β, δ), that is, quasi hyperbolic preferences, where it is possible to model the sharp drop in utility between periods leading to individual self-control problem[20]. The difference, compared to the standard model, lies in the presence of this parameter $\beta \leq 1$ representing the present bias, "*capturing the main finding of the experiments* (DellaVigna, 2009, p. 318)". However, with $\beta = 1$ this is the standard case, where the model is only influenced by a constant discount factor $\delta < 1$ between periods, whichever they are. In this model, an individual incurs a cost at the time of undertaking the task ($t = 1$), and receives his payoff delayed one period ($t = 2$).

[19] However, it should be noted that in the work of DellaVigna and Pope (2018), no evidence was found strong enough to claim that the quasi-hyperbolic discount correctly illustrated the time preference of individuals.

[20] Notwithstanding the almost generalized use of quasi-hyperbolic preferences, Ainslie (2012) argues that to represent self-control the most correct would be the "pure" hyperbolic preferences.

3.1.2. Rational decision-making paradigm

Given the above described new feature of this individual, one of the doubts that arises is how far individuals' rationality extends. Although this agent is not the *Homo Economicus* concerning information processing, he will continue to be a sophisticated decision maker. But, within economics, in what way is it possible to relate defense of beliefs (seemingly irrational behavior), and still allow individuals to be rational?

On the one hand, intertemporal utility maximization holds individuals within the framework of rational behavior (standard rational decision making) where they pursue self-interest according to their image of their ability[21]. Individuals want the best for themselves at every moment of time, in such a way that maximization must be seen as "*a good first approximation to what instinct, education or learning will ultimately produce* (Tirole, 2002, p. 640)". It is clear that agents make mistakes, especially by the use of heuristics that in certain contexts lead to seemingly irrational decisions, but if I do not try to construct an analysis of the cases where the heuristic is correctly used (on average), otherwise it is likely that I focus only on the exceptions.

On the other hand, one of the points of potential escape from the standard rational scenario could be the limited attention incorporated in this model, but this use of only a subset of information is framed within the rationality of the agent. As agents will not be considered naive decision makers, they will be aware that when there are persistent incentives to manipulate or suppress negative news they suffer from a systematic tendency to remember (or collect) only good news (Tirole, 2002). For that reason, they will have a kind of metacognition (introspection[22]) through which they identify if they are successively suppressing the bad news, which allows them to discount the non-existence of negative news. Likewise, if there are third-party incentives, they will not take them in a naive way once they

[21] However, this behavior may not be enough to prevent the emergence of self-traps.

[22] For example, more simply, this insight may represent an individual's ability to perceive that in the past he may have tended not to pay attention to information that might contradict his beliefs. As when, in a political discussion, agents argue successively in their favor in such a way that they are not even available to understand the contrary argument. What is expected is that the individual who does so, at a later introspective moment, realizes that he embarked on such a strategy that allowed him to defend his beliefs.

can infer their ability through these third-party behaviors. This metacognition will be represented in this model by a Bayes rule[23] (Bénabou, 2015).

3.2. Ex-ante Belief Manipulation Strategies: Learn or Avoid?

What is the best information structure to maximize utility, when an individual is time inconsistent, and only observes information about his ability in the period in which he has to undertake a project? The answer lies on the a priori belief about his ability, and on the news, he may receive about his true ability. This individual can choose two different information structures in period $t = 0$: avoid or learn information in period $t = 1$.

3.2.1. Model Formal Features

This model is built fundamentally on the value of motivation that an individual derives from his self-confidence and is based fundamentally on the works of Carrillo and Mariotti (2000) and Bénabou and Tirole (2002a). This motivation part in the motivated beliefs model addresses the incentives to avoid or receive information with the aim of defending long-term goals. Thus, understanding properly why an individual searches for self-confidence (explained earlier in topic 2.3.1) will help explaining the reason for belief defense. I start by considering that in this model the individual is risk neutral and that his relevant time horizon is three periods $(t = 0, 1, 2)$. An individual chooses in period $t = 0$ (Self 0) the information structure that his reincarnation will have in period $t = 1$ (Self 1)[24]. This structure choice is extremely relevant because it determines the amount of available information that Self 1 has. Information may be transmitted from different information sources: through information request/feedback (e.g.: a test result, performed at $t = 0$ and received at $t = 1$) or through experience (e.g.: learning by doing) (Bénabou & Tirole, 2002a). Note that I will only consider two extreme cases:

- When an individual in $t = 1$ learns as much information as possible: all the information that Self 0 knows and all information that can be learned or observed

[23] Bénabou and Tirole, in their various works, argue that even considering less sophisticated agents, "*as long as they are not too naïve* (Bénabou & Tirole, 2002b, p. 876)" results will be similar.

[24] It is worth bearing in that, in the model represent below in this thesis, Self 0 and Self 1 will represent the incarnations of the same individual at different periods of time. However, it could also represent communication between different individuals.

between periods (such as the result of a medical examination performed at t = 0 or a new agricultural production technique learned in a training course, which will evaluate its good or bad ability for future production).

- When he does not learn nothing Self 0 did not know. That is, Self 1 only knows the probability of his ability being good or bad. Making his decision depend on his priori belief about his ability.

In period $t = 1$ this individual decides if he should undertake a project ($e \in \{0,1\}$). If he chooses to undertake it, he incurs in a cost $c > 0$ (disutility cost). The project has an associated expected success probability[25], represented by θ, which can result in a benefit V to this individual in period $t = 2$.

Additionally, this model incorporates a quasi-hyperbolic discount factor β (with $\beta < 1$), which represents individual dynamic inconsistency (or "weakness of will"), in addition to a standard discount factor δ ($\delta < 1$) . This is due to a "present bias" empirically verified and that results in a problem of self-control, which means that discount rates are much lower at short horizons than at mote distant ones[26]. Finally, consider that u_t represents the individual payoff in period t and that $E_t(u_{t+1})$ represent the individual expectation at period t about utility in period $t + 1$.

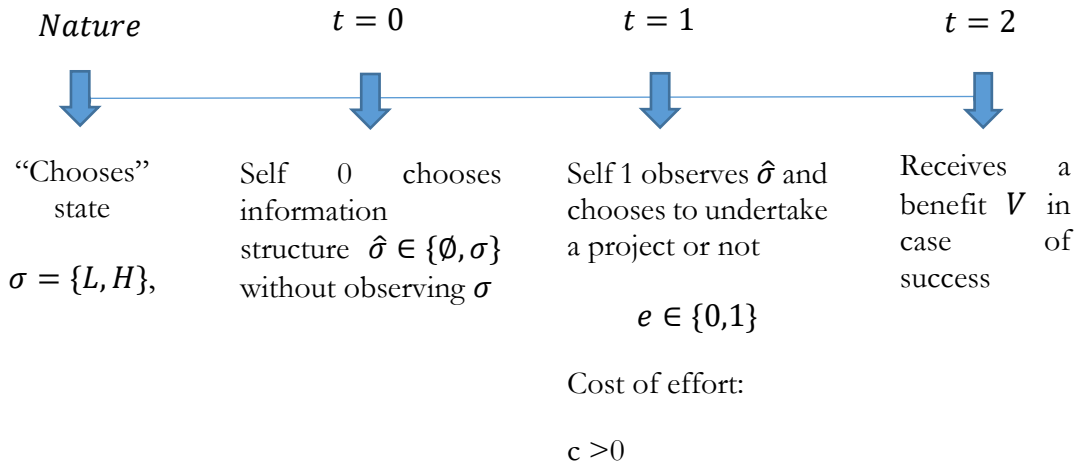


Figure 1 – Ex-ante Belief Manipulation Strategies

[25] This success probability may be also seen as individual ability.
[26] This issue was previously discussed in section 3.1.1.

From the viewpoint of Self 0, the expected payoff is given by

$$U_0 = \beta[\delta E_0(u_1) + \delta^2 E_0(u_2)]$$

which may be described as follows:

$$\begin{cases} 0 & \text{if } e = 0 \\ \begin{cases} \beta\delta(-c + \delta V) & \text{if } e = 1 \text{ and the project is successful} \\ \beta\delta(-c) & \text{if } e = 1 \text{ and the project is unsuccessful.} \end{cases} \end{cases}$$

Thus, if the project is undertaken, the expected payoff simplifies to

$$\beta\delta[-c + \delta\theta_0 V].$$

In this way, from Self's 0 point of view it is preferable to undertake the project if, and only if, the payoff of undertaking the effort is greater than the payoff of not undertaking

$$\beta\delta[-c + \delta\theta_0 V] \geq 0 \;\Rightarrow \theta_0 \geq \frac{c}{\delta V}.$$

From Self 1 viewpoint, as he is influenced by time inconsistency, his payoff is given by

$$U_1 = u_1 + \beta\delta[E_1(u_2)],$$

and in which, Self 1 payoff may be described as:

$$\begin{cases} 0 & \text{if } e = 0 \\ \begin{cases} -c + \beta\delta V & \text{if } e = 1 \text{ and the project is successful} \\ -c & \text{if } e = 1 \text{ and the project is unsuccessful.} \end{cases} \end{cases}$$

Then, in case of investment, his expected payoff is equal to

$$-c + \beta\delta\theta_1 V.$$

Thus, Self's 1 he will undertake the project if and only if

$$-c + \beta\delta\theta_1 V \geq 0 \Rightarrow \theta_1 \geq \frac{c}{\beta\delta V}$$

and this is the fundamental requirement for the individual to undertake the project since the decision is made by the Self 1.

Nevertheless, from this simple exposition it is possible to verify that an individual incurs an intrapersonal conflict to decide whether or not to undertake the project. This is because the expected return, given the dynamic inconsistency, has to be higher so that from the standpoint of Self 1 he chooses to undertake it. When an individual analyzes, at $t = 0$, if he invests in a project in the period $t = 1$, he will decide to undertake the project if the expected return is $\theta_0 \geq \frac{c}{\delta V}$, but when this individual approaches this period, given his preference for a reward in the present (present bias), it requires a higher expected return ($\theta_1 \geq \frac{c}{\beta\delta V}$) for him to undertake the project.

3.2.2. Avoiding vs Learning Information

Finally, regarding the formation of beliefs, it is important to understand the impact that the choice of information structure has. Let's first focus on the case where individual's ability is unknown by Self 1, that is, he just knows that nature "chose" the news about individuals ability $\sigma = \{L, H\}$. With probability $q \in [0,1]$ ability is high (H), and with probability $(1 - q)$ ability is low (L). In addition, it should be noted that, $\theta_H > \theta_0 > \theta_L$. Good news always gives the individual better news than his prior belief, and that negative news always provides worse news than his prior belief. Thus, Self 0 has a prior belief given by the mean of the probability distribution function $F_0(\theta)$ and he chooses to remain uninformed in period $t = 1$. Since the information that an individual has between periods does not change, his posterior belief about his ability will be also given by the mean of the probability distribution function $F_1(\theta)$[27], wich is $\theta_0 = q\theta_H + (1 - q)\theta_L$.

[27] As defended by Bénabou and Tirole "*with risk neutrality the mean* $\bar{\theta} \equiv \int_0^1 \theta dF(\theta)$ *will be a sufficient static for* $F(\theta)$ (Bénabou & Tirole, 2002a p. 878)". Thus, as $F_1(\theta) = F_0(\theta)$, hence $\theta_1 = q\theta_H + (1 - q)\theta_L = \theta_0$.

In contrast, an individual can get to know his true ability: the true value of θ, which can be θ_H or θ_L. In this approach, for simplicity, the individual learns his ability at $t = 1$ without costs, by conducting examinations or requests for feedback in $t = 0$. As state above, this ability depends on the nature or on the "state of the world" $\sigma = \{L, H\}$. So, an individual no longer depends on his prior belief about his ability $F_0(\theta)$, since he will have at $t = 1$ an accurate posterior belief of what his ability really is.

3.2.3. Four distinctive cases

I will present here four different cases, in which the decision to receive or avoid information about individual ability will vary according to the prior belief and with the possibility of new information increase or decrease individual's belief about his success probability. Additionally, it is necessary to bear in mind that the subsequent analysis is performed from the standpoint of Self 0, since it is the time when an individual decides whether to inform Self 0 or not. But referring, where necessary, to the conflict of interests between Self 0 and Self 1, since it is Self 1 that will undertake the project or not.

Case 1: $\theta_L \geq \frac{c}{\beta\delta V}$

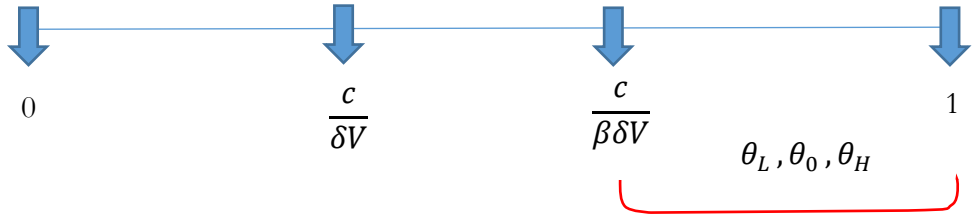


Figure 2 – Case 1: Self 1 always invests (no conflict of interest)

This is a case in which an individual believes that his ability (without concrete information) is sufficiently high for the project to always be successful ($\theta_0 > \frac{c}{\beta\delta V}$). Thus, this individual will always undertake the project, if he does not receive information. It is trivial to say that if this individual decides to receive such information, it will not affect his decision in the period $t = 1$. Even if he receives bad news, given that $\theta_L \geq \frac{c}{\beta\delta V}$, it is easily understandable that with this news Self 1 wishes to undertake the project. The same is true, without information, for Self 0. As such there is no conflict of interest between temporal selves.

Case 2: $\frac{c}{\delta V} < \theta_L < \frac{c}{\beta\delta V} < \theta_0$

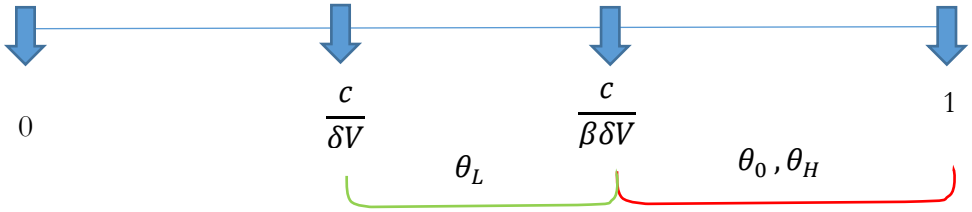


Figure 3 – Case 2: Self 1 does not invest if news are bad (conflict of interest)

Here, I'm dealing with a more interesting case where there is a conflict of interest. As before, this individual has a good prior belief ($\theta_0 > \frac{c}{\beta\delta V}$) about his ability, which allows him to undertake a project if he does not receive news about his ability. However, if he receives information, and this is less ($\frac{c}{\delta V} < \theta_L < \frac{c}{\beta\delta V}$) than what an individual considers necessary at the time of the decision to undertake the project, it is certain that he will procrastinate. However, Self 0 continues to consider that it would be optimal to invest in the project (since for him it is optimal to invest if $\theta_0 > \frac{c}{\delta V}$, but Self 1 decides, after receiving the bad news, that his skill is so low that he will not undertake the project. Regarding the importance of time inconsistency, it must be said that in this case the greater the time inconsistency (lower β) the greater the gain from maintaining Self 1 uninformed. As Self 1 is more time inconsistent, if informed of his true ability, the loss due to procrastination is increasing $[\frac{c}{\delta V}, \frac{c}{\beta\delta V}]$.

Case 3: $\theta_L < \frac{c}{\delta V} < \frac{c}{\beta\delta V} < \theta_0$

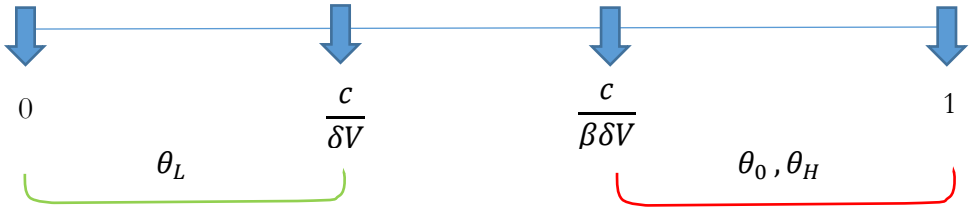


Figure 4 – Case 3: Self 1 does not invest if news are bad (no conflict of interest)

Consider now that this individual has a good prior belief ($\theta_0 > \frac{c}{\beta\delta V}$) about his ability, which allows him to undertake a project if he does not receive news of his ability. However, if an individual find that his ability is much lower ($\theta_L < \frac{c}{\delta V}$) than the ability he considered

to have a priori, $\theta_0 > \frac{c}{\beta\delta V}$, then the transmission of information will be very useful to avoid a situation of overconfidence. In this situation it is clearly preferable to inform Self 1 that his ability is much lower than he imagined he had, and it is easy to understand that both Self 0 and Self 1 do not wish to undertake the project in this situation.

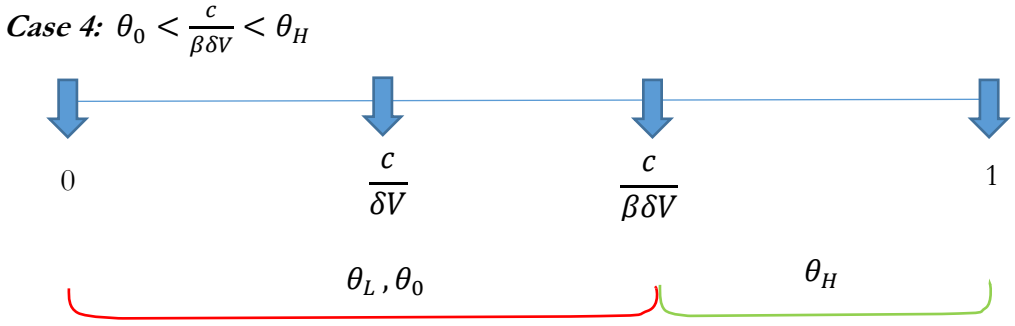


Figure 5 – Case 4: Self 1 only invests if news are good (possible conflict of interest)

Consider that Self 1 believes that, without information, his ability is too low ($\theta_0 < \frac{c}{\beta\delta V}$) to undertake the project. However, he can be in two different situations: does Self 0 believed he has enough ability for the project $\frac{c}{\delta V} < \theta_0 < \frac{c}{\beta\delta V}$ or does Self 0 thought that he has practically no skill for the project ($\theta_0 < \frac{c}{\delta V}$)? In none of them does Self 1 want to undertake the project (albeit with $\frac{c}{\delta V} < \theta_1 < \frac{c}{\beta\delta V}$ Self 0 wants). Nevertheless, if Self 1 discover that its true ability is $\theta_H > \frac{c}{\beta\delta V}$, there is here some possible gain from having better information since it would help recovering Self's 1 motivation to undertake a project. Here, there is no conflict of interest if good news is received, once both selves will agree that the best decision is to undertake the project.

3.2.4. Learn or avoid information?

As the amount of information available to Self 1 is defined by Self 0, it makes sense to assess which of the situations will be most beneficial intertemporally. To do so, Self 0 should evaluate what is the strongest motive (confidence maintenance or overconfidence) thus verifying whether information has a positive or negative value for this individual. If the risk of overconfidence is greater, then there will be an information gain from making available additional information for Self 1. Otherwise, for the confidence maintenance motive, there will be a loss for having more information, and for this reason Self 0 will prefer not to provide

additional information (Bénabou & Tirole, 2002a). Formally, comparing from Self 0 viewpoint, the expected payoff that will be obtained whether or not additional information on individual ability is conveyed, will help him to understand which his best option is.

Regarding ***Case 2 and Case 3***, consider that Self 1 does not have information because Self 0 avoided the transfer of information. Thus, from Self 0 viewpoint (U_0^A) his payoff will be:

$$U_0^A = \beta\delta(-c) + \beta\delta^2[q\theta_H + (1-q)\theta_L]V$$

$$U_0^A = \beta\delta\{(-c) + \delta[q\theta_H + (1-q)\theta_L]V\}$$

When Self 1 learns information, from Self 0 viewpoint (U_0^R) his payoff will be:

$$U_0^L = (1-q)[0] + q[\beta\delta(-c) + \beta\delta^2\theta_H V]$$

$$U_0^L = \beta\delta[q(-c) + \delta\theta_H V]$$[28]

It should be noted that this analysis is valid only for the cases mentioned above in which Self 1 invests if it does not receive information and does not invest when he receives information and it is negative. Thus, comparing the payoffs which represent the most interesting case from Self 0 viewpoint

$$U_0^A > U_0^L$$

$$\beta\delta\{-c + \delta[q\theta_H + (1-q)\theta_L]V\} > \beta\delta\{q[(-c) + \delta\theta_H V]\}$$

$$-c + (1-q)\delta\theta_L V > q(-c)$$

$$\theta_L > \frac{c}{\delta V}$$

it is possible to conclude that this individual in period $t = 0$ wants to prevent the transfer of information whenever Self 1 ability is above $\theta_L > \frac{c}{\delta V}$. That is, he wants to prevent Self 1

[28] Thus, this payoff only reflects the case in which the individual undertakes effort $[(1-q)\theta_H]$, and as such gets a payoff.

from being informed whenever information about his ability does not indicate that undertaking the project is a costly mistake.

Focusing now on ***Case 4***, noting that this analysis is valid when an individual does not invest if he does not receive information and that he invests when he receives positive information. From the point of view of Self 0, if he avoids transferring information (U_0^A) he will receive a payoff

$$U_0^A = 0.$$

If, Self 1 receives information, from Self 0 viewpoint (U_0^L) his payoff will be

$$U_0^L = q[\beta\delta(-c) + \beta\delta^2\theta_H V]$$

Without proceeding with mathematical determination, it is easily discernible that with

$$U_0^A > U_0^L$$

$$0 > q[\beta\delta(-c) + \beta\delta^2\theta_H V]$$

an individual always prefers to receive information, otherwise his payoff will be null.

Finally, ***Case 1***, in which an individual invests whether he has information (positive or negative) about his ability, whether he has no news. Thus, from Self 0 viewpoint, without giving new information to Self 1, his expected payoff (U_0^A) will be

$$U_0^A = \beta\delta\{(-c) + \delta[q\theta_H + (1-q)\theta_L]V\}.$$

Receiving information, Self 0 expected payoff (U_0^L)will be

$$U_0^L = (1-q)[\beta\delta(-c) + \beta\delta^2\theta_L V] + q[\beta\delta(-c) + \beta\delta^2\theta_H V].$$

In this case, it is also not necessary to advance to the mathematical determination of the critical value, since it is easily noticeable that the two payoffs are equal. For this individual, it is indifferent to choose the structure of the information because regardless of the news the project will be successful.

3.3. Ex-post Belief Manipulation Strategies: Register or Forget?

Another important side of this Motivated Beliefs model is the possibility for individuals to affect the probability of remembering data. With this addition to the model it will be possible to relate the two aspects of self-deception: motivation (functional reason for wanting to defend beliefs) and cognition (the way they actually do it) (Bénabou, 2015; Bénabou & Tirole, 2002a). Simply put, this distinct approach allows individual's awareness to be selective, memory imperfect and attention limited.

3.3.1. Additional Formal Features

I will continue to consider a risk neutral individual that has a relevant time horizon of three periods $(t = 0, 1, 2)$. This individual will receive information about is ability in period $t = 0$, where he has to decide whether to manipulate his memory or not, which will thus influence the news Self 1 will have. The issue of information manipulation will be very important because the decision of undertaking a project or not will be influenced by how an individual evaluates the reliability of his memory. However, as in the previous section, I will consider only two extreme cases:

- When the individual fully relies on his memory. This means that in the period $t = 1$, an individual has all the information that was actually received, since this individual is unable to manipulate his memory.
- When an individual does not trust at all in his memory because whenever he has an opportunity he "hides" negative news. Thus, in period $t = 1$ the individual does not trust anything of what his memory "transmits" to him, because in period $t = 0$ this individual tried to forget all the negative news that has received. For example, he sought only news that corroborated his beliefs, or ignored (or paid less attention to) negative news.

As before, it is Self 1 who decides to undertake the project or not ($e \in \{0,1\}$). Choosing to undertake the project, an agent incurs in a cost ($c > 0$). The dynamic inconsistency factor β (with $\beta < 1$) will also be present in addition to the standard rational discount factor δ ($\delta < 1$). The project has an associated expected success probability, represented by θ, which can result in a benefit V to this individual in period $t = 2$.

The first significant change is to transform this model into a "bad news model" (meaning that in this model, only the reception of negative news will be evaluated). So, I will consider that Nature "decides" that with probability q individuals do not receive any news ($\sigma = \emptyset$), and that with probability $(1-q)$ he receive negative news ($\sigma = L$). When an agent does not receive bad news in the initial period these cannot be "manufactured ex nihilo (Bénabou & Tirole, 2002a, p. 889)". This allows an individual to doubt his beliefs only when he does not have information in period $t = 1$.

Thus, depending on the observed signal, Self's 0 expected ability will be described as

$$\begin{cases} E_0[\theta|\sigma = L] = \theta_L \\ E_0[\theta|\sigma = \emptyset] = \theta_H \end{cases} \qquad \theta_H > \theta_L$$

However, if this information is threatening to individual's self-confidence, he may have interest in repressing it.

Let

$$\lambda = \Pr[\hat{\sigma} = L|\sigma = L]$$

denote the probability that negative news will be present in memory in period $t = 1$. Note that, agents can increase or decrease this probability. So, in period $t = 1$ an agent recalls information received in the previous period as $\hat{\sigma} \in \{\emptyset, L\}$. Aware of these incentives to distort information when news is negative ($\sigma = L$), an individual evaluates his memory "reliability" of having no news $\sigma = \emptyset$ through a Bayes rule

$$r \equiv \Pr[\sigma = \emptyset \,|\hat{\sigma} = \emptyset, \lambda] = \frac{q}{q + (1-q)(1-\lambda)}.$$

Consequently, Self 1 expected ability is given by

$$\begin{cases} E_1[\theta|\hat{\sigma} = L] = \theta_L \\ E_1[\theta|\hat{\sigma} = \emptyset] = \theta(r) = r\theta_H + (1-r)\,\theta_L. \end{cases} \qquad \theta(r) > \theta_L$$

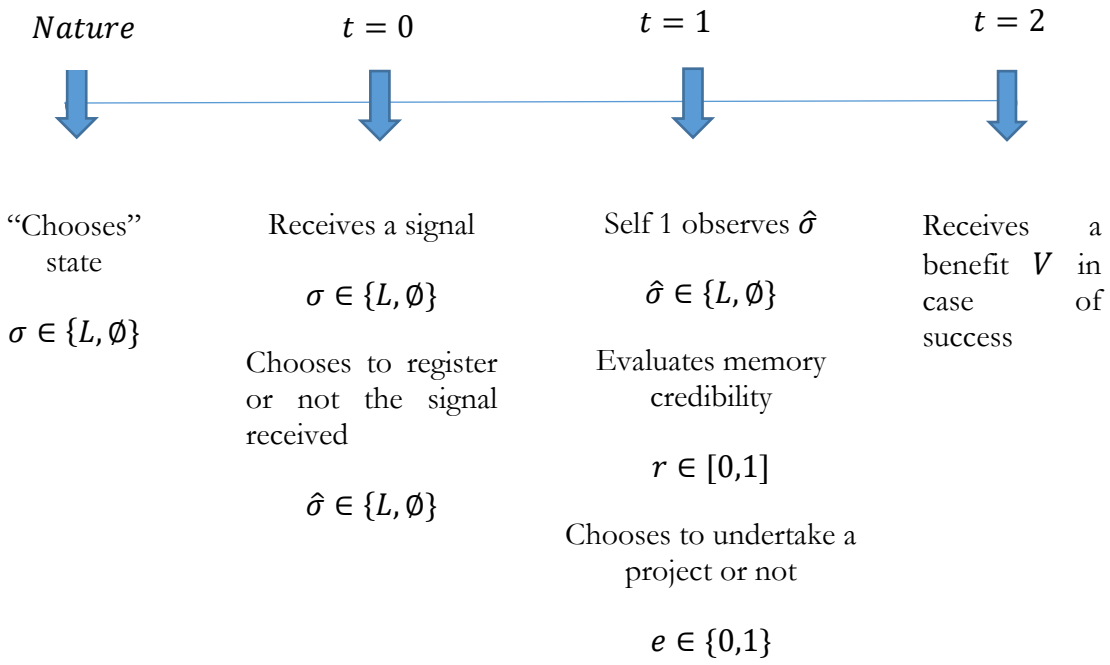


Figure 6 – Ex-post Belief Manipulation Strategies

Finally, recall that U_t represents the intertemporal utility evaluated in period t, that u_t represents the individual payoff in period t and that $E_t(u_{t+1})$ represent the individual expectation at period t about utility in period $t+1$.

Thus, as before, from Self 0 viewpoint, the expected payoff is given by

$$U_0 = \beta[\delta E_0(u_1) + \delta^2 E_0(u_2)]$$

which implies that for the Self 0 the project should be undertaken if, and only if

$$\beta\delta[-c + \delta\theta_0 V] \geq 0 \;\Rightarrow \theta_0 \geq \frac{c}{\delta V}.$$

From Self 1 viewpoint, as he is influenced by time inconsistency, his expected payoff is given by

$$U_1 = u_1 + \beta\delta[E_1(u_2)]\,,$$

then from Self 1 viewpoint it is preferable to undertake the project if, and only if,

$$-c + \beta\delta\theta_1 V \geq 0 \;\Rightarrow \theta_1 \geq \frac{c}{\beta\delta V}.$$

This is also the critical value for an individual to decide to undertake the project, as it is in period $t = 1$ that the decision is made.

3.3.2. Not Registering vs Registering Negative News

Regarding the way individuals defend their beliefs, it is worth understanding how the manipulation of received news (especially negative ones) is a very important factor. On the one hand, we have the case where an individual decides not to register negative news. I consider here a case in which Self 0 repress totally bad news (corresponding to a probability of remembering negative news of $1 - \lambda = 1$ or $\lambda = 0$) in order to motivate himself to undertake the project in additional cases where he would not do it if he recalls that news. This difficulty in registering negative news correctly comes when an individual compares his expected return, conditional to his behavior at the moment of decision making, and finds out that his preferences are dynamically inconsistent. Still, sometimes an individual in period $t = 1$ may not believe in the "no news" report as he knows his tendency not to register negative news in period $t = 0$. In order to decide what to do about projects, this individual will have to remember what his expectation before the arrival of any news would be.

On the other hand, I consider that this individual is unable to manipulate his memory. In this case Self 0 correctly makes reports about the received news. As before, Self 1 will invest in a project if he has not received negative news. Nevertheless, in this new situation Self 0 reports are absolutely truthful ($\lambda = 1$) and Self 1 reliability in Self's 0 reports about having (or not) received any negative news is total ($r = 1$). As a consequence of this, it is certain that this individual will have different time preferences when he faces negative news (since he prefers to undertake a project from $t = 0$ viewpoint and does not want to do it from $t = 1$ viewpoint), which lead to a conflict of interest.

3.3.3. Two Distinctive Cases

I will present here formally the two distinct cases presented above, in which the individuals decision will depend on the news that they receive, as well as on the reliability of their memory. In both cases, I will consider these two hypotheses:

- From Self 0 viewpoint it is always preferable to invest in a project.

$$\frac{c}{\delta V} < \theta_L < \theta_H$$

- From Self 1 viewpoint, he invests just in case he has not received any news.

$$\theta_0 < \frac{c}{\beta\delta V} < \theta_H$$

Case 1: Individuals cannot commit to registering negative news

Here is analyzed a scenario where an individual cannot commit to registering negative news because he has a tendency to manipulate his memory. Thus, in the period t = 1, he will not remember any negative news, so he would think he did not receive them. But, is it possible to imagine an equilibrium where Self 1 would invest if he has any negative news?

Consider then that

$$\lambda = \Pr[\hat{\sigma} = L | \sigma = L] = 0$$

since it is never optimal to register bad news because when in period $t = 1$, this individual will never undertake a task (as $\theta_L < \frac{c}{\beta\delta V}$) and that is not what he prefers when he receives the news (from Self 0 viewpoint $\frac{c}{\delta V} < \theta_L$).

Consequently,

$$r = \frac{q}{q + (1-q)(1-\lambda)} = q$$

Thus, as he totally represses bad news, he understands that when he is in period $t = 0$ he will not register any negative news (because they make him procrastinate). But because of this he also does not trust his memory, in such a way that this individual has a posterior belief about his ability that is equal to his prior belief. Thus, as I defined earlier $E_1[\theta | \hat{\sigma} = \emptyset] = \theta(r) = r\theta_H + (1-r)\,\theta_L$, and considering the extreme case in which an individual in period $t = 1$ does not rely at all in his memory, this individual will have a success probability given by $\theta(r) = q\theta_H + (1-q)\,\theta_L$ (equal to his ex-ante success probability $\theta_0 = q.\theta_H + (1-q)\theta_L$).

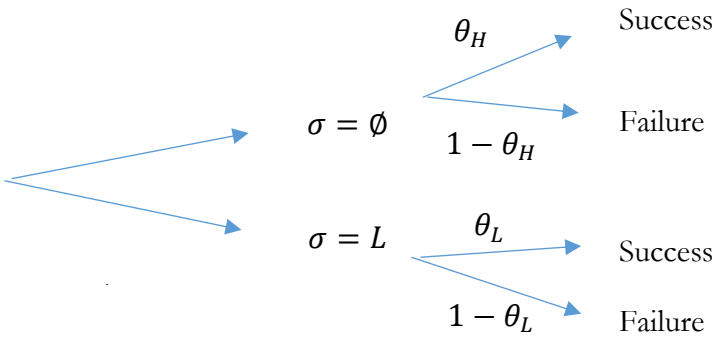


Figure 7 – Ex-ante success probability

This means that since $\theta_0 < \frac{c}{\beta\delta V}$ Self 1 will not undertake a project when he is able to manipulate his memory. Once an individual does not trust his memory, he is in a situation similar to the absence of information. Thus, it is possible to conclude that this individual ability to manipulate his memory, in period $t = 0$, backfires. Contrary to what he would prefer from an ex-ante $(t = 0)$ viewpoint, his future incarnation never undertakes a project, even when the "no news" report is truthful. So, there is a conflict of interest between temporal selves.

As shown below, when it is possible to manipulate memory, Self 0 wants to invest whatever the news and Self 1 does not want to invest whatever the news:

$$\theta_{\sigma\in\{L,\emptyset\}} > \frac{c}{\delta V} \rightarrow \text{From Self 0 viewpoint prefers } e = 1$$

$$\theta_{\hat{\sigma}\in\{L,\emptyset\}} < \frac{c}{\beta\delta V} \rightarrow \text{From Self 1 viewpoint prefers } e = 0$$

Thus, the expected utility of Self 0 when it does not register the news (forget them - U_0^F) is given by

$$U_0^F = 0.$$

Summarizing, in equilibrium, when an individual does not have any news at period $t = 1$, he never invests. So that, he never invests, whatever the news, when he can manipulate his memory.

Case 2: Individuals are unable to manipulate the news

Now, consider a scenario where an individual is unable to manipulate the news he receives. However, abserve that:

- If $\sigma = \emptyset$ there is no decision to be taken.
- If $\sigma = L$ an individual has to decide whether or not to record the negative information.

Therefore, when bad news is received, the payoff to register is never higher than not registering the negative news (because he will be slightly better than in the extreme case - there will be some credibility of his memory). So not registering bad news is a weakly dominant strategy.

Thus, in the period t = 1, he will always correctly remember the news he has received. Let's evaluate again the following question: is it possible to imagine an equilibrium where Self 1 would invest if he has any negative news?

Consider then that

$$\lambda = \Pr[\hat{\sigma} = L | \sigma = L] = 1$$

since he always registers bad news in period $t = 0$, this individual will never undertake the project (as $\theta_L < \frac{c}{\beta\delta V}$) when they arise. As before, that is not what Self 0 prefers when he receives the news (from Self 0 viewpoint $\frac{c}{\delta V} < \theta_L$).

Consequently,

$$r = \frac{q}{q + (1-q)(1-\lambda)} = 1$$

thus, since this individual always correctly records the negative news, he totally relies on his memory as he is not able to manipulate it. For this individual, it is only preferable to undertake a project since he does not receive any news $E_1[\theta|\hat{\sigma} = \emptyset] = \theta_H$ because with good news Self 1 knows that $\frac{c}{\beta\delta V} < \theta_H$. With negative ones $E_1[\theta|\hat{\sigma} = L] = \theta_L$ and Self 1

knows that $\theta_L < \frac{c}{\beta\delta V}$ so he does not invest. Therefore, Self 1 trusts totally his memory, in such a way that is posterior belief is a correct one and updated by a Bayes rule.

It is thus understood that when this individual is unable to manipulate his memory, in the absence of information, Self 1 will undertake the project as Self 0 would want and as such there is no conflict of interest. This does not happen if negative news is received, since Self 0 would like to undertake the project, but Self 1 does not undertake it. Thus, it is possible to conclude that without manipulating his memory, this individual can at least undertake a project when $\hat{\sigma} = \emptyset$.

As shown below, without being possible to manipulate memory, Self 0 wants to invest whatever the news and Self 0 just wants to invest without news:

$\theta_{\sigma\in\{L,\emptyset\}} > \frac{c}{\delta V}$ → From Self 0 viewpoint prefers $e = 1$

$\theta_{\hat{\sigma}=\emptyset} < \frac{c}{\beta\delta V}$ → From Self 1 viewpoint prefers $e = 0$

$\theta_{\hat{\sigma}=L} > \frac{c}{\beta\delta V}$ → From Self 1 viewpoint prefers $e = 1$

Thus, the expected utility of Self 0 when he registers all the news (remember them at $t = 1$: U_0^R) is given by

$$U_0^R = q[\beta\delta(-c) + \beta\delta^2\theta_H V]$$

Thus, it is verified that exists an equilibrium in which Self 1 invests if he has not negative news but only when Self 0 is not able to manipulate the news he receives. When comparing the payoff, it is easily verified that between the two situations, both selves have a bigger payoff in this second example, where they cannot manipulate memory.

4. Conclusion

Despite the fact that most of the work on decision-making in economics is based on very rigid hypotheses about human behavior, the psychological enrichment present in this approach demonstrates that it is not a problem for the formality of the economy, since it allows to mimic behaviors that would not otherwise be represented. As objectified, evidence was found that individuals were able, within their rationality, to manipulate what they believe. The multidisciplinary literature on the subject has no doubt in affirming that individual beliefs are in fact motivated, and to a large extent by their instrumental use for motivation. Furthermore, it was found that motivated reasoning strategies that lead individuals to engage in ex-ante and ex-post belief-manipulation strategies are mostly beneficial, from an instrumental point of view, however in special cases this may constitute not a benefit but a self-trap.

Thus, beliefs are important, and have to be an issue to be analyzed when characterizing the behavior of individuals. Since it is not a direct consequence of the information individuals receive, it is always fundamental to understand the motivations that lead an individual to maintain a particular belief. Often the purpose of maintaining beliefs is their ability to motivate particular behaviors, but the reasons for defending these beliefs are much more complex. The maintenance of self-esteem, self-confidence or reputation are some of them.

Instrumentally, individuals are interested in maintaining their beliefs whenever they allow them to undertake a project. When they do not have the opportunity to know the information immediately, they are often dependent on future preferences (different from the current ones because they are dynamically inconsistent) of their reincarnation in the following period, when the investment decision must be made. This is not always a bad thing, since it is verified that if the prior belief in the individual's ability is incorrectly too high, this dependence on his reincarnation will result in the effort procrastination, which will be the right decision. On the other hand, whenever an initial belief in ability is low, an individual depends on the information he will receive to regain his motivation, otherwise he will not undertake the project. The role that dynamic inconsistency has in motivated beliefs should be emphasized, since it is in cases where ability is high enough for Self 0 and too low for Self

1 that this problem reveals the conflicts of interest that may arise between temporal selves. The gains and losses that can result from the decisions about the information structure are all the greater the greater the dynamic inconsistency.

When it was later considered that the memory about the information received could be manipulated, there are two perfectly distinct cases to be treated. When positive news is received, individuals would rather not have the ability to manipulate their memory, as this may lead to disbelief. When individuals receive negative news, manipulation of beliefs leads them to a situation where manipulation is a weakly dominant strategy, so individuals may benefit from it. In this analysis, the model was converted into a "bad news model". Also, what the explicited cases demonstrate, first, is that manipulation of beliefs is helpful to an individual, provided that it does not fall into the extreme case of permanent self-deception in which the payoff of self-deceiving is lower than that of recalling all memories correctly. Thus, it is concluded that the manipulation allows to motivate to undertake a project in some cases, but it can not be said that such manipulation does not lead to a self-trap.

4.1. Future work

The conceptual framework presented here can still be extended, and as such there is room for future work. Perhaps the attempt to frame all those that are considered non-standard beliefs such as overconfidence, the law of small numbers, and the bias of projection, may be a direction to follow. Of the three possibilities mentioned above, overconfidence is already possible to fit into this model. Of the remainder, the projection bias seems to be the one that is more ambitious to fit into this framework of motivated beliefs. This is due to the need to also consider the consumption value (and not just the instrumental value) in the formalization of the model.

Logically, improvement from the conceptual framework depends not only on the addition of new biases. A very interesting change, which can lead to very interesting results, would be the relaxation of the hypothesis of rationality. Following the direction that Bénabou and Tirole presented in their work of 2002 (Bénabou & Tirole, 2002a), rationality could be parameterized in order to explore its influence on the beliefs of individuals. This would transform the individuals present here in imperfect Bayesian ones, insofar as relaxation of

this hypothesis could allow individuals not only to be unaware of what they may have forgotten, but also to forget that they have forgotten information previously received.

5. References

Abell, G. O., & Greenspan, B. (1979). Human births and the phase of the moon. *The New England Journal of Medicine*, *300*(2), 96.

Ainslie, G. (1975). Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control. *Psychological Bulletin*, *82*(4), 463–296.

Ainslie, G. (1985). Beyond microeconomics. Conflict among interests in a multiple self as a determinant of value. In J. Elster (Ed.), *The Multiple Self* (pp. 133–175). Cambridge University Press.

Ainslie, G. (2012). "Pure hyperbolic discount curves: eyes open self-control." *Theory and Decision*, *73*, 3–34.

Ainslie, G. (2016). "The Cardinal Anomalies that Led to Behavioral Economics: Cognitive or Motivational?" *Managerial and Decision Economics*, *37*, 261–273.

Ainslie, G., & Herrnstein, R. J. (1981). Preference reversal and delayed reinforcement. *Animal Learning & Behavior*, *9*(4), 476–482.

Battaglini, M., Bénabou, R., & Tirole, J. (2005). "Self-control in peer groups." *Journal of Economic Theory*, *123*(2), 105–134.

Bem, D. J. (1972). Self-Perception Theory. *Advances in Experimental Social Psychology*, *6*, 1–62.

Bénabou, R. (2013). "Groupthink: Collective delusions in organizations and markets." *Review of Economic Studies*, *80*(2), 429–462.

Bénabou, R. (2015). "The Economics of Motivated Beliefs." *Revue d'économie Politique*, *125*(5), 665–685.

Bénabou, R., & Tirole, J. (2002a). "Self-confidence and personal motivation." *Quarterly Journal of Economics*, *117*(3), 871–915.

Bénabou, R., & Tirole, J. (2002b). "Self-Knowledge and Self- Regulation : An Economic Approach." In *The Psychology of Economic Decisions: Volume One: Rationality and Well-Being* (pp. 137–167).

Bénabou, R., & Tirole, J. (2003). "Intrinsic and extrinsic motivation." *Review of Economic Studies*, *70*(3), 489–520.

Bénabou, R., & Tirole, J. (2004). "Willpower and Personal Rules." *Journal of Political Economy*, *112*(4), 848–887.

Bénabou, R., & Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *Quarterly Journal of Economics*, *126*(2), 805–855.

Bénabou, R., & Tirole, J. (2016). "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives*, *30*(3), 141–164.

Benoît, J.-P., & Dubra, J. (2011). "Apparent Overconfidence." *Econometrica*, *79*(5), 1591–1625.

Brown, J. D., Dutton, K. A., & Cook, K. E. (2001). "From the top down: Self-esteem and self-evaluation." *Cognition and Emotion*, *15*(5), 615–631.

Caplin, A., & Leahy, J. (2001). "Psychological Expected Utility Theory and Anticipatory Feelings." *Quarterly Journal of Economics*, *116*(1), 55–79.

Carrillo, J. D., & Mariotti, T. (2000). "Strategic Ignorance as a Self-Disciplining Device." *The Review of Economic Studies*, *67*(3), 529–544.

Chung, S.-H., & Herrnstein, R. J. (1967). Choice and Delay of Reinforcement. *Journal of the Experimental Analysis of Behavior*, *10*(1), 67–74.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature*, *47*(2), 315–372.

Dellavigna, S., & Malmendier, U. (2004). Contract Design and Self-Control: Theory and Evidence. *The Quarterly Journal of Economics*, *119*(2), 353–402.

DellaVigna, S., & Pope, D. (2018). What Motivates Effort? Evidence and Expert Forecasts. *Review of Economic Studies*, *85*(2), 1029–1069.

Eil, D., & Rao, J. M. (2011). "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics*, *3*(2), 114–138.

Fernando Pessoa. (1968). *Textos Filosóficos. Vol. II Fernando Pessoa (Estabelecidos e prefaciados por*

António de Pina Coelho.) (Ática). Lisboa.

Gabaix, X., & Laibson, D. (2005). "Bounded Rationality and Directed Cognition." *NBER Working Paper No. 11755*.

Gilovich, T. (1991). *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*. New York: Free Press.

Greenwald, A. G. (1980). "The Totalitarian Ego Fabrication and Revision of Personal History." *American Psychologist*, *35*(7), 603–618.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 407–424.

Kahneman, D. (2011). *Thinking, Fast and Slow. New York: Farrar, Straus and Giroux.*

Kahneman, D., & Tversky, A. (1979). "Prospect Theory : An Analysis of Decision under Risk." *Econometrica*, *47*(2), 263–292.

Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R., & Dolan, R. J. (2014). "Depression is related to an absence of optimistically biased belief updating about future life events." *Psychological Medicine*, *44*(3), 579–592.

Koszegi, B. (2010). "Utility from anticipation and personal equilibrium." *Economic Theory*, *44*(3), 415–444.

Köszegi, B. (2006). "Ego Utility, Overconfidence and Task Choice." *Journal of the European Economic Association*, *26*(5), 673–1.

Kruger, J., & Dunning, D. (1999). "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments." *Journal of Personality and Social Psychology*, *77*(6), 121–1134.

Laibson, D. (1997). "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, *112*(2), 443–478.

Loewenstein, G. (1996). "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, *65*(3), 272–292.

Logan, F. A. (1965). Decision making by rats: Delay versus amount of reward. *Journal of Comparative and Physiological Psychology*, *59*(1), 1–12.

Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of honest People: A theory of Self-Concept Maintenance. *Journal of Marketing Research*, *XLV*, 633–644.

Mischel, W., & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, *16*(2), 329–337.

Navarick, D. J., & Fantino, E. (1976). Self-control and general models of choice. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*(1), 75–87.

O 'Donoghue, T., & Rabin, M. (1999). Doing It Now or Later. *American Economic Review*, *89*(1), 103–124.

Phelps, E. S., & Pollak, R. A. (1968). On Second-Best National Saving and Game-Equilibrium Growth. *The Review of Economic Studies*, *35*(2), 185–199.

Puri, M., & Robinson, D. T. (2007). "Optimism and economic choice." *Journal of Financial Economics*, *86*(1), 71–99.

Rabin, M. (1998). "Psichology and Economics." *Journal of Economic Literature*, *36*(1), 11–46.

Rabin, M. (2002). "A perspective on psychology and economics." *European Economic Review*, *46*(4–5), 657–685.

Rachlin, H., & Green, L. (1972). Commitment, choice and self-control. *Journal of the Experimental Analysis of Behavior*, *17*(1), 15–22.

Sayette, M. A., Loewenstein, G., Griffin, K. M., & Black, J. J. (2008). Exploring the Cold-to-Hot Empathy Gap in Smokers. *Psychological Science*, *19*(9), 926–932.

Sharot, T., & Garrett, N. (2016). "Forming Beliefs: Why Valence Matters." *Trends in Cognitive Science*, *20*(1), 25–33.

Shybut, J. (1968). Delay of gratification and severity of psychological disturbances among hospitalized psychiatric patients. *Journal of Consulting and Clinical Psychology*, *32*(4), 462–468.

Simon, H. A. (1955). "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*, *69*(1), 99–118.

Solnick, J. V., Kannenberg, C. H., Eckerman, D. A., & Waller, M. B. (1980). An experimental analysis of impulsivity and impulse control in humans. *Learning and Motivation*, *11*(1),

61–77.

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *Source: The Review of Economic Studies*, *23*(3), 165–180.

Svenson, O. (1981). Are we all less risky and more skillfull than our fellow drivers? *Acta Psychologica*, *47*, 143–148.

Swann, W. B., Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist*, *62*(2), 84–94.

Thaler, R. (1980). "Toward a positive theory of consumer choice." *Journal of Economic Behavior & Organization*, *1*(1), 39–60.

Thaler, R. (1981). "Some empirical evidence on dynamic inconsistency." *Economics Letters*, *8*, 201–207.

Thaler, R. (2000). "From Homo economicus to Homo sapiens." *Journal of Economic Perspectives*, *14*(1), 133–141.

Thaler, R. (2016). "Behavioral Economics : Past , Present , Future." *American Economic Review*, *106*(7), 1577–1600.

Thaler, R., Tversky, A., Kahneman, D., & Schwartz, A. (1997). "The Effect of Myopia And Loss Aversion On Risk Taking: An experimental test." *Quarterly Journal of Economics*, *12*(2), 647–661.

Tirole, J. (2002). "Rational irrationality: Some economics of self-management." *European Economic Review*, *46*(4–5), 633–655.

Tversky, A., & Kahneman, D. (1974). "Judgment under uncertainty: Heuristics and biases." *Science*, *185*(4157), 1124–1131.

Weinstein, N. D. (1980). Unrealistic Optimism About Future Life Events. *Journal of Personality and Social Psychology*, *39*(5), 806–820.