Journal of Institutional Economics (2026), **22**, e9, 1–19
doi:10.1017/S1744137426100411

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Rules and temptations: which came first?

Elias L. Khalil

School of Economics, Administration and Public Policy, Doha Institute for Graduate Studies, Doha, Qatar
Email: elias.lafi.khalil@gmail.com

(Received 8 July 2025; revised 30 December 2025; accepted 2 January 2026)

Abstract

For received theories, (suboptimal) temptations arise first, and, consequently, people set up rules or institutions to control them. Hence, any deviation from institutions is suboptimal. However, these received theories face an anomaly, coined here the 'Holiday License Paradox': Why would people who adopt optimal institutions turn around and designate 'holidays' (cheat days) that allow them to indulge in suboptimal consumption? To solve this paradox, this paper reverses the entry point: people first set up rules – whereas temptations are identifiable only with respect to those rules. This solution raises a new question: what is the origin of rules? People adopt rules to control 'temerity', i.e., overconfidence. This raises a further question: what is the origin of temerity? Temerity is a default heuristic expressing the optimal response in life-and-death decisions. Thus, temerity-as-heuristics is rather efficient on average. However, temerity can become excessive, and, at second approximation, people adopt rules to control temerity. Once we regard rules or institutions to come first, i.e., prior to temptations, it becomes possible to solve the Holiday License Paradox.

Keywords: dual process theory; heuristics; holiday licences (cheat days); present-bias preferences; principal-agent framework; puritan lifestyle; self-deception; temerity (overconfidence)

'Pride goeth before destruction, and an haughty spirit before a fall'

Proverbs 16:18–King James Version (KJV)

Introduction

The research question

This paper proposes an unorthodox thesis. It disputes the supposition of received theories. Received theories include present-bias approaches (e.g., Strotz, 1955; Laibson, 1997; O'Donoghue & Rabin, 1999), the multiple-self framework (Fudenberg & Levine, 2006; Thaler & Shefrin, 1981), and the dynamic consistency theory (Gul & Pesendorfer, 2001, 2004a, 2004b). Despite their differences, received theories basically suppose that temptations are the primitives while the decision maker (DM) adopts rules to control them. This paper proposes the contrary: rules rules come first, while temptations can be defined only *vis-à-vis* rules. This unorthodox thesis raises the question about the origin of rules. One major part of this paper is devoted to answering this question.

The proposed thesis basically rearranges two basic phenomena (temptations and rules) that must be the fundamental elements of any theory of temptations (weakness of will). The research question concerns the entry point:

© The Author(s), 2026. Published by Cambridge University Press on behalf of Millennium Economics Ltd. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

Research Question: Should a theorist commence with rules to explain temptations? Or should she follow received theory, i.e., commence with temptations as the primitives and consequently explain rules as the attempt to control those primitives?

The proposed entry point, i.e., rules, allows researchers to provide endogenous accounts of temptations. Rules can be regarded as a kind of institution or constitutional design. There are two kinds of such rules or constitutional designs. The first are rules or constitutional designs that involve individual behaviour *vis-à-vis future self*: the DM designs rules where the aroused temptations concern over-consumption (i.e., under-saving). The over-consumption undermines the wellbeing of the future self. The second kind are rules, or, more commonly recognised, constitutional designs that involve individual behaviour *vis-à-vis contemporaneous others*: the DM designs rules where the aroused temptations concern also over-consumption. However, the over-consumption here undermines the wellbeing of pertinent others, i.e., cooperators that the DM has made promises to.

As explained elsewhere (Khalil, 2020), both acts of over-consumption are analytically identical. Although this thesis is not the subject of this paper, it can be a source of confusion. The confusion is expected given that the claim that both kinds of over-consumption are identical is contrary to textbook economics, where it models intertemporal allocation (present self *vis-à-vis* future self) differently from how it models the prisoner's dilemma game (present *self vis-à-vis* contemporaneous others). While it models the DM in intertemporal allocation as concerned about the wellbeing of their future self, it models the DM in the prisoner's dilemma as concerned only about the wellbeing of their own self. As explained below, there is a mechanism of punishment of cheaters in intertemporal allocation that is similar to the mechanism in the prisoner's dilemma game. If we are convinced that the two acts of over-consumption or cheating are isomorphic, it should not be surprising that the proposed theory, which focuses mostly on over-consumption that undermines the wellbeing of the future self, should also be a theory regarding over-consumption that undermines the wellbeing of pertinent contemporaneous others.

The proposed claim of isomorphism is not new. A few theorists and philosophers (e.g., Elster, 2000, pp. 88–174; Teubner & Beckers, 2013; Waluchow & Kyritsis, 2023) have noted that the set of constitutional rules regarding individual liberties is the condition that DMs, theoretically behind the veil of ignorance, demand prior to becoming members of the respective political community. The demand for such constitutional rules entails that, irrespective of the faction that controls the office of the executive or the legislature, no faction can squash rights that are *ex ante* dear to all.

Indeed, this paper uses the principal-agent framework, relying on the isomorphism thesis, namely, reneging on promises made to contemporaneous others amounts to succumbing to temptations. That is, reneging is analytically identical to failing to live to promises made to the future self.

This paper focuses only on temptations, a major phenomenon that has captured the interests of many behavioural economists (see Camerer, 2003; Camerer et al., 2004; Dhami, 2017, 2025). There are many other important phenomena that have also captured the attention of behavioural economists. However, they fall outside the scope of this paper. These phenomena include habits (e.g., heuristics and routines), fairness (e.g., the ultimatum game), loss aversion, the income-happiness paradox (Easterlin paradox), the endowment effect, and framing effects. This range does not include deviations from rationality arising from anxiety, such as procrastination, the certainty effect (Allais paradox), and the ambiguity effect (Ellsberg paradox). Indeed, it is hard to keep track of all these phenomena, not to mention putting them into an efficient taxonomy (see Khalil, 2025a, 2025b).

Definitions: Temptations, rules, institutions, constitution

When the DM succumbs to temptations, the DM succumbs to *weakness of will*. The consequent consumption is by definition suboptimal. Such consumption is usually over-consumption, i.e., where the DM's weakness of will undermines the wellbeing of their future self (or the wellbeing of pertinent

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

contemporaneous others). The consumption can also be suboptimal in the opposite direction, namely, where the DM succumbs to *stiffness of will*. This paper does not discuss under-consumption, i.-e., choices motivated by stiffness of will.

It suffices, however, to highlight two examples of stiffness of will: (i) miserly behaviour in the case of intertemporal allocation where the future self is the focus (see Khalil, 2017a); and (ii) sentimental foolishness exhibited in suboptimal acts of helping others in the cases of cooperation where pertinent others are the focus (see Khalil, 1999).

To illustrate temptations in the sense of succumbing to over-consumption, let us imagine a model of a single DM who is planning in the present, at t_1, consumption and saving for fifty periods. In period t_{50}, she expects to die at the end and, hence, plans to bequeath the invested capital to her daughter. She cares about the daughter's wellbeing with the same intensity as she cares about her own wellbeing at t_1. In this model, the DM has well-defined preferences regarding present and future consumption, a certain capital at t_1 that equals, say, \$1.1 million, a uniform discount rate of future consumption that equals, say, 10%, and a constant investment rate of return per period that equals the discount rate. The DM has neither an income from labour nor from rental property. At the beginning of t_1, the DM decides to consume \$100,000 and to save \$1 million. At the beginning of t_2, she receives the investment return (\$100,000) and consumes it while keeping the \$1 million invested. Indeed, the optimal consumption in each period is \$100,000. At t_{50}, she dies at the end of the period, as she expected. The daughter inherits the \$1 million. Assuming that the daughter has the same preferences and budget constraint, she would also choose the optimal consumption bundle, \$100,000 at t_{51} and all following periods, assuming that she also has one offspring.

Temptation is defined here as the deviation of the DM (or her daughter) from the optimal consumption plan, \$100,000 per period. It is usually the case to define weakness of will as succumbing to 'temptation' of consuming any amount at any period that is *above* \$100,000. However, it is analytically correct, arising from 'stiffness of will' (Khalil, 2017a), to consider any consumption *below* \$100,000 to be also succumbing to 'temptation'.

Regarding rules, they amount to informal and formal guidelines at the level of the DM as well as the level of the collective. At the collective level, they are cultural norms, government regulations, laws, and rights placed in constitutions. Such rights usually pertain to limiting the supremacy of one actor, subjecting the declarations of war, the raising of taxes, and other major policies to the careful scrutiny of other actors within the state.

As such, the term 'rules' denotes in this paper one kind of institution, namely, the self-commanded restraints that decision makers (DMs) impose on themselves à la Adam Smith's (1976) concept of 'self-command' (see Khalil, 2010). This paper does not study 'institutions' in the sense of habits and laws designed to define and regulate property relationships, precautions against uncertainty, moral limits of contracts, the functions of the judiciary, or the duties of the state in encouraging saving, innovations, and economic development. These types of institutions have been studied by Williamson (1985), North (1990, 1991), and Hodgson (2015; see Khalil, 2012, 2013). The focus of this paper is narrower than the focus of these authors.

Once we study these authors and find the judicious balance between privately and publicly owned property and further specify the informal norms, laws, and regulations associated with such property relationships, what is the source of succumbing to temptations? Why do people tend to violate the institutional foundations of the economic society? Do the rules, in the sense of self-commanded restraints, arise in light of preferences conceived as primitives – or *vice versa*?

The paper simply does not study what determines the judicious institutions à la Williamson, North, or Hodgson. It takes them as givens. On the basis of these givens, it asks a question corresponding to Smith's self-command: Do we have self-commanded restraints (rules) because of human predilection toward tantalising preferences, or *vice versa*?

The scope of this paper is even more narrow. It does not discuss two variants of the mentioned self-commanded restraints. The first variant that this paper ignores is 'general rules' in Smith's (1976: 156–161) sense, namely, as principles that DMs impose on themselves to avoid

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

self-deception. Self-deception, following the definition of Smith, is the technology that DMs employ to masquerade their deviations from the self-commanded restraints (see Khalil, 2009, 2016, 2017b; Khalil and Vasudevan, 2025). The second variant that this paper ignores is formal 'general rules', i.e., regulations and laws. This paper does not focus on when informal 'general rules', i.e., 'norms', become formal 'general rules'. This analysis involves a research question outside the scope of this paper.

Received theories mentioned at the outset also ignore 'general rules' – the principles that help DMs avoid self-deception – as this paper does. These theories study 'rules', i.e., self-commanded restraints to control temptations as technologies designed to amend human partiality toward present consumption.[1] Despite their differences, they effectively conceive temptations as primitives, as 'present-bias preferences'.[2]

The holiday license paradox

For received theories, present-bias preferences entail that DMs choose a consumption of suboptimal goods only because they are situated in the present, i.e., they would not choose them excessively if they were situated in any period in the future.

To state the received theories succinctly:

> **Received Theories:** DMs suffer from present-bias preferences that pose temptations; in turn, DMs adopt self-imposed rules, ranging from commitments to so-called 'pre-commitments', to prevent succumbing to temptations. That is:
>
> **Temptations → Rules**

This paper offers a radical proposal: we need to reverse the order. At the first moment, DMs adopt self-imposed rules. What they feel as temptations appears only at the second moment, only in reference to the self-imposed rules. That is, received theories simply place the proverbial cart in front of the horse.

The motivation behind this radical proposal is to solve a contradiction – an anomaly that is not noted in the received literature. This paper coins it the 'Holiday License Paradox':

> **The Holiday License Paradox:** DMs erect constraints to ward off stand-alone consumptions from, say, the tempting Y set, i.e., those deemed suboptimal. So, *ceteris paribus,* bundles of the Y set are *always* 'bads', i.e., suboptimal. Failing to save or exercise is always 'bad'. On the other hand, in order to enhance wellbeing, DMs erect institutions that license limited choices of bundles from the Y set. How come, given the *ceteris paribus* condition, DMs consider the Y set to be always suboptimal but allow choices from the Y set as exceptions, like holiday licences?

Holiday licences allow DMs to splurge during vacations, e.g., to overeat, under-exercise, and so on. Holiday licences are simply licences to cheat, but cheat within a boundary. DMs observe holiday licences in many cultures. Indeed, holiday licences are evident in almost all societies, contemporaneous and historical. From a quick survey of the sociological and anthropological literature (e.g., Bakhtin, 1965; Dietler and Hayden, 2001; Falassi, 1987; Goody, 1982), a researcher would be surprised to find stable or everlasting puritan societies that have managed to totally suppress indulgence that violates the rules.

[1]Stated briefly (see Appendix 2), the self-imposed rules to combat temptations closely correspond to what George Ainslie (2001) calls 'bright lines', what Adam Smith dubs 'general rules' (Khalil, 2009, 2010; Smith, 1976: 156–161), what Gul and Pesendorfer (2001) name 'self-control', and what Bénabou and Tirole (2004) call 'personal constraints'. The rules can be external, in which case they are called 'commands' (Schelling, 1984), 'menu' choice (Gul and Pesendorfer, 2001), 'constitutions' (Elster, 2000), and 'nudges' (Thaler and Sunstein, 2008).

[2]Instead of the term 'present-bias preferences', some philosophical literature uses the term 'cyclical preferences' in order to characterise them as the violation of the transitivity axiom (e.g., Andreou, 2023, Ch. 3).

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

There are many versions, forms, and explanations of events that remotely or closely resemble the holiday licence phenomenon that this paper identifies. This paper cannot discuss the phenomenon in all its aspects and various forms (see Khalil, 2026c). At its most elemental level, the holiday licence phenomenon is paradoxical. Given that circumstances have not changed, why would humans consume in holidays bundles of goods that are deemed to be bad in all instances?

Indeed, the proposed key thesis of this paper is an attempt to solve the holiday license paradox. To solve the paradox, we need to rearrange the two basic elements (temptations and rules) in contra-distinction to the arrangement in received theories of weakness of will. The identified paradox stems from engaging theory with history. It emerges from engaging rational choice theory and optimal rules, on the one hand, with the historical record of most societies insofar as they set up holiday licences while knowing that such licences undermine the optimal rules, on the other hand.

This paper's methodology of engaging theory with history can be traced back to Ibn Khaldûn. For example, Ibn Khaldûn (1967, vol. 2: 103–111) maintained that the prince or the sultan should avoid predatory public finance: a reasonable and stable tax rate can ensure greater public revenue than a capricious and high tax rate. He enlisted his historical accounts of the debacles of many political dynasties in history in light of a theory of public finance that modern economists characterise as 'predatory' (e.g., Olson, 2000; see Khalil, 2007a, 2007b).

Similar to Ibn Khaldûn's methodology, this paper uses holiday licences in historical and modern societies to highlight a dilemma that faces standard, neoclassical rational choice theory insofar as it commences with tempting preferences. For such a theory, the decision strategy is clear: at each point, the DM should abide by rules (pre-commitments) to avoid splurging, overeating, and under-exercising. That is, at first approximation, the DM should lead a puritan lifestyle (e.g., Fitouchi et al., 2023). However, evidently, DMs embrace holiday licences, i.e., violating the rational puritan lifestyle.

Let us grant that the received starting point faces the holiday license paradox. Still, why does reversing the entry point, i.e., starting with rules, promise to solve it so easily? Once this paper uncovers below the origin or source of the adopted rules, it becomes easy to realise that such origin allows the DM to design rules and, simultaneously, design with ease the exceptions, i.e., the holiday licences.

As Table 1 sums up, the radical proposal, called below *core thesis I,* faces a question: if rules are not the products of tempting preferences, what is the origin of rules? In response, this paper advances *core thesis II*: the origin of rules is temerity in the sense of overconfidence – defined as the tendency to form unfounded beliefs about the states of nature (environment) to rationalise reckless behaviour.[3] However, in turn, core thesis II faces another question: what is the origin of temerity? In response, this paper advances core thesis III: the origin of temerity is heuristics, i.e., deep-seated machinery of temerity-guided beliefs needed in life-and-death decisions. However, heuristics may generate excessive temerity, prompting DMs to adopt rules to restrain it.

The paper discusses, in three sections, core thesis I, core thesis II, and core thesis III, respectively. It proceeds to identify the basic motive for the adoption of rules in a preliminary model. The last section concludes. This paper's contribution resides in the three core theses and, further, the preliminary model that explains the problem with temerity.

A separate online link (https://eliaskhalil.com/download/Appendices-Rules-and-Temptations.pdf) contains three pertinent appendices. Appendix 1 distinguishes temptations from addictions, on the one hand, and the certainty effect, on the other hand. Appendix 2 offers a critical review of different theories of temptations, including rational addiction theory, dynamic inconsistency, the multiple-self approach, dynamic consistency, and dual process theory. Appendix 3 lists three major payoffs of the proposed theory.

[3]Another paper (Khalil, 2026b) elaborates on the key concept 'temerity'. It broadens it to denote optimal heuristics resulting from the asymmetry of capability improvement. When the DM expects the benefits from confrontation to equal the benefits from non-confrontation (withdrawal), the DM chooses the confrontation strategy because, as can be easily assumed, it allows for greater capability improvement than the withdrawal.

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

Table 1. Questions, received answers and proposed answers

Questions	Answers of received theories	Proposed answers	Proposed lexicon
Question 1: What is the Origin of Temptations?	Temptations are preferences and, as such, they are the primitives	Answer 1 – Start with Rules: Rules → Temptations	Core Thesis 1 – The Rules-Dependent Temptations Hypothesis
Question 2: What is the Origin of Rules?	Rational Decision Makers design rules to manage temptations	Answer 2 – Start with Temerity: Temerity → Rules	Core Thesis 2 – The Temerity-Dependent Rules Hypothesis
Question 3: What is the Origin of Temerity?	The origin of temerity can be of different sources – notwithstanding that temerity is unrelated to temptations	Answer 3 – Start with Heuristics: Heuristics → Temerity	Core Thesis 3 – The Heuristics-Dependent Temerity Hypothesis

Question 1: What is the Origin of Temptations?

The Answer: Start with Rules

The radical proposal, called here core thesis I, can be expressed tersely:

Core thesis I (*The Rules-Dependent Temptations Hypothesis*): Tempting preferences cannot be the entry point. The entry point should be the rules, the set or restraints. That is,

Rules → Temptations

The core thesis I, 'the rules-dependent temptations hypothesis', entails that tempting preferences can be identified only in reference to rules, where such rules form a part of the set of constraints. That is, tempting preferences cannot stand-alone preferences. They originate from the already set-up rules.

The core thesis entails that the world of temptations is *monist*: there is a single phenomenon called 'temptations', and, hence, it behoves us to develop a single theory: either temptations are primordial while rules appear later or, as this paper argues, rules appear prior to temptations. However, what if the world of temptations is *non-monist* – what if we have two different phenomena of temptations? The first, call it the 'Ten Commandments' phenomenon of the Old Testament, is the one investigated in this paper: there are rules that exist for some reason, and only against such rules can we identify temptations. The second, call it the 'Ulysses' phenomenon, where Ulysses tries to fight off the tempting songs of the Sirens (Elster, 2000), is the supposition of an intense taste for songs or chocolate that acts as the primordial entity. This supposition means that the DM sets up rules to command excessive consumption.

However, given that we need to define 'excessive consumption', the Ulysses phenomenon inescapably reverts to the Ten Commandments phenomenon. Let us start with a DM who has an intense taste for chocolate. Still, the DM has to decide what is the optimum normal consumption – above which would be 'excessive consumption'. Let us say, given the DM's set of preference, the optimum is 50 grams per day. Let us say that there is a special sale for one day where the optimum for the DM under focus becomes 70 grams per day. The DM finds that if he, given the special sale, consumes 70 grams, he cannot stop and actually consumes 150 grams. To prevent such suboptimal consumption, the DM sets up a 50-gram-per-day rule that makes him disregard special circumstances such as special sales. The desire to consume more, as in the case of special sales, would be 'a tempting option'.

Thus, to define a 'tempting option', we must already have in place what is the optimum or, what is more common, the rule that expresses the optimum in average (normal) circumstances.

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

A defence of rules

How should we reconcile the rational decision strategy of leading a puritan lifestyle with the actuality of holiday licences that are anti-puritan? We could reconcile rationality with holiday licences if we can theorise that when people design rules, they also can easily design exceptions, i.e., holiday licences. This would be easy to theorise only if we suppose that rules precede temptations. However, this leads to Question 2.

Question 2: What is the Origin of Rules?

The Answer: Start with Temerity

Given core thesis I, what is the origin of rules? This paper postulates that DMs design rules to regulate temerity:

Core thesis II (*The Temerity-Dependent Rules Hypothesis*): DMs adopt the self-imposed rules in order to manage temerity. That is,

Temerity → Rules

Core thesis II, 'the temerity-dependent rules hypothesis', traces rules to temerity. This means that rules can be identified only in reference to temerity.

This paper defines 'temerity's the cavalier attitude that lowers one's reading of the risks regarding non-auspicious outcomes – i.e., increases the probabilities of fortunate states of nature – below what the facts suggest. Such a cavalier attitude is the basis of the formation of unfounded beliefs about the states of nature. Such an attitude leads the DM to take actions that she would not have taken if she had read the facts objectively and, consequently, held fact-based prior beliefs about the states of nature. Examples of temerity include reckless driving based on downplaying the likelihood of accidents, overeating based on dismissing the likelihood of ill health, and failing to take extra precautions on sea voyages based on lowering the risk of storms.

In contrast, we may reserve the term 'prudence' to denote an attitude that generates properly formulated risk assessments, leading to well-founded beliefs.[4]

Once we see temerity, especially excessive temerity, rather than tempting preferences, at the origin of rules, we can solve the holiday licence paradox as promised. If we conceive rules as a function of excessive temerity, the DM can modify the rules during holidays, i.e., allow for licences to over-consume, *without worry*. After the holiday, the DM can return to the rule, i.e., avoid overconsumption. The holiday acts as a known limit for the overconsumption and, hence, preserves the integrity of the rules. Put plainly, the DM tells the self, 'I can splurge and take risks, but this activity is limited to holidays and vacations'. In this manner, the episode of overconsumption will not distort the prudent rule, i.e., will not lead to cavalier or temerarious beliefs about the environment.

In contrast, if we travel along the received theories – i.e., conceive rules as a function of present-bias preferences – the DM should not modify the rules: at every instance the DM should avoid the suboptimal choice. As an example, the rule 'do not under-save' must be true in every instance – except when circumstances change as in the case of the emergency of saving a life. The holiday licence is far from an emergency to save a life. The DM cannot find it useful to violate the rule in some instances. Thus, it is difficult for received theories to explain holiday licences.

[4]The term 'attitude' is foreign to standard economists, as it does not easily fit in a toolbox consisting of information, beliefs, preferences, and constraints. Following the psychological literature (e.g., Maio et al., 2018), attitudes are deep-seated behavioural patterns that can be innate or learned. They amount to a routinised way of employing seemingly fixed traits such as limbs, peaks, language, and so on, to attain a goal. Thus, 'attitude' parallels the use of the term 'habit' in either the sense of 'routine' or 'heuristics'.

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

In response, received theories may resort to *ad hoc* qualifications – which effectively weaken their explanatory apparatus (Khalil, 1989). One qualification is the reasoning that holiday licences are 'exceptions' to habits. As one may break away from a habit under an extraordinary circumstance, one may assign a holiday as an exception. Such reasoning regards the continuous avoidance of temptations to be burdensome. When the DM adopts a rule of abstemiousness, she also adopts an exception, i.e., a holiday that allows the self to 'sin'. When the DM yields to a temptation, it allows the DM to release a repressed desire that otherwise can wreak havoc.

In this explanation, the rule, with its respective holiday licence, is on average optimal. However, this explanation portrays the rule regarding temptations as a 'habit'. While the rule might become habitual, it does not mean its origin (insofar as it is about temptations) is the same as the origin of habits. Habits, as the term is defined here, are not about rules to combat temptations. Habits include heuristics that humans form as first impressions or generalisations. Habits also include routines such as taking the same route home or brushing the teeth. As argued in Appendix 2 (see also Khalil, 2022; Khalil and Amin, 2023), DMs adopt habits (ranging from heuristics to routines) to minimise the cognitive and physiological cost of deliberation of each instance.

Indeed, dual process theory explains habits as the manifestation of mental economy (e.g., Grayot, 2020; Kahneman, 2011). This theory proposes that decisions or choices proceed along either a fast process, called 'System 1', or a slow process, named 'System 2'. This is not the paper to discuss the origin of the two systems–not to mention the basis of the switching between them (Khalil, 2025c).

It is sufficient to state that the basis that gives rise to habits differs from the basis that originates rules regarding temptations. (Notwithstanding, as shown below, this paper relies on dual process theory to explain temerity, which is another phenomenon). We cannot explain the holiday licence phenomenon as 'exceptions' to habits – given that rules regarding temptations are not related to the difference between System 1 and System 2 (dual process theory).

To recapitulate, with core thesis II, the cheat days or holiday licences phenomenon is no longer an anomaly. The phenomenon is only an anomaly in reference to received theories – i.e., theories that commence with suboptimal, present-bias preferences.

A defence of temerity

To recapitulate, DMs adopt rules to regulate temerity, which this paper coins the 'temerity-dependent rules hypothesis'. That is, rules do not stand alone. They can only be defined in reference to temerity.

This paper postulates 'temerity' as the origin of rules, as it is the most convincing origin that can *easily* explain, besides rules, the exceptions to the rules, i.e., holiday licences. That is, the specified origin is a successful hypothesis if it carried the two tasks at first approximation, i.e., without *ad hoc* assumptions and tortured qualifications.

Temerity is a cavalier *attitude* that leads the DM to exaggerate the likelihood of auspicious outcomes. This exaggeration takes place without the DM being necessarily conscious of his misguiding attitude. Such an attitude can be expressed in the asymmetrical reaction to outcomes. For example, let us take two outcomes that are the products of two choices, where each choice requires the same effort. However, the outcomes differ: one is successful while the other is not. The difference is the result of two exogenous shocks of equal probability. If the outcome is successful, the DM attributes the success more to skill/effort than to luck. If the outcome is unsuccessful, the same DM attributes the failure more to luck than to skill/effort. We can conclude that these DMs have a temerarious attitude, as they tend to inflate their ability – not realising that ability is not in question.

Let us take a game of die tossing, where the die is fair. The DM gains $1 when she tosses an even number and loses $1 when she tosses an odd number. The expected value is zero. A risk-neutral DM would be ready to pay no more than zero to participate in this game. However, if the risk-neutral DM is conceited, i.e., operating on the belief that 'I am special' or 'I feel lucky today', she would be ready to pay

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

a positive amount, say $10 for 100 rounds, based on the belief that, ex ante, there will be at least 60 times out of the 100 rounds where the toss generates an even number.[5]

In everyday examples, if the DM was lucky in guessing the outcome of a roulette spin, he would upgrade the likelihood of an outcome in the following spin, thinking that 'it is my lucky day'. However, if the DM was unlucky in the first round, he would not symmetrically downgrade the likelihood of an outcome in the following spin. Temerarious DMs seem eager to build a causal story where it neglects that outcomes, following Galton's law, regress to the average (see Galton, 1889; Mulligan, 1999).

In his light, temerity is the basis of unfounded beliefs in a specific sense:

> **Temerity:** When DMs over-recognise the role of luck when outcomes are negative while under-recognising the role of luck when outcomes are positive, such DMs operate according to a cavalier attitude, temerity, i.e., a belief that presents the likelihood of success higher than justified by ex ante evidence.

In contradistinction to temerity, when DMs are motivated by prudence, they uphold objective beliefs about the likelihood of success, i.e., beliefs supported by available information.

We can use the terms 'arrogance' and 'smugness' as equivalent to 'temerity'. However, we should not use the term 'optimism' as equivalent to 'temerity' – contrary to how some authors use the term (e.g., Sharot, 2011). Optimism has the connotation of being motivated by a positive outlook toward life, ever thankful and ever hopeful, even in dire catastrophes and circumstances (Khalil, 2023). Optimism suggests transcendence, i.e., going beyond everyday material constraints, which temerity does not suggest.

We may use the terms 'overconfidence' or 'over-estimation' as equivalent to 'temerity'. However, these terms suggest narrower scope than 'temerity'. They limit the issue to the assessment of ability, while 'temerity' suggests both assessment of ability and beliefs about the states of nature.

Specifically, as Moore and Schatz (2017) clarify, while the term 'overconfidence' has three separate meanings in the literature, they all about the assessment of ability. The first meaning is the 'over-estimation' of an ability, as in solving word puzzles; the second is the 'overplacement' of one's ability *vis-à-vis* the ability of a group, as in automobile driving; and the third is the 'overprecision' in one's belief that she is correct.

In any case, economists generally do not distinguish overconfidence in the narrow sense of assessment of ability, on the one hand, and temerity in the broad sense that includes beliefs about the states of nature, on the other hand. For example, Bhaskar and Thomas (2019) detect temerarious attitude by measuring the extent that DMs exaggerate their ability.[6]

The task of separating temerity from overestimation of ability is difficult. As noted above, the DM is usually not cognisant of her own temerity (see Barber and Odean, 2001; Kőszegi, 2006). This is similar to self-deception: temerity (as well as self-deception) would vanish once the DM recognises the attitude as such. Probably the DM senses his temerity as if it is about ability rather than the mistaken exaggeration of luck.

Question 3: What is the Origin of Temerity?

The Answer: Start with Heuristics

[5]The special feeling 'it is my lucky day' seems to resemble the tendency to form patterns of clouds and stars where none exists. This is known as the 'hot-hand fallacy'. (Note, however, the original data that inspired the 'hot-hand fallacy' term support the intuition that there is a real pattern in the winning streak of basketball games, as Miller and Sanjurjo (2018) show).

[6]Examine also Buehler et al. (1994), Camerer and Lovallo (1999), Dunning et al. (1989), Hoelzl and Rustichini (2005), Kadous et al. (2014); McDonald and Ross (1999), Swann and Gill (1997), Tetlock (2005), and Thaler (2016).

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

Given core thesis II, what is the origin of temerity? This paper postulates that DMs adopt temerity as a survival strategy:

Core thesis III (*The Heuristics-Dependent Temerity Hypothesis*): DMs adopt temerity as a habit in the sense of a heuristic to enhance survival in critical situations. That is,

Heuristics → Temerity

Core thesis III, 'the heuristics-dependent temerity hypothesis', stipulates that, ultimately, the entry point is heuristics.

This paper maintains that temerity is an optimal heuristic in the game of life. DMs take risks not only in everyday decisions about how to drive, what to buy, and whether to accept a job offer. DMs also have to take risks that have definite outcomes with regard to staying alive. When DMs face critical choices in life-and-death situations, they face binary outcomes: staying alive or dying. Consequently, DMs find it better to err on the side of recklessness, i.e., temerity, than to err on the side of the opposite, i.e., prudence in the sense of well-founded caution (see Khalil, 2026b).

A defence of heuristics

Temerity-as-heuristics are on average optimum. That is, heuristics do not always engender the correct responses. A heuristic-based action might produce, *ex post*, a negative outcome. The optimum heuristics are destined on many occasions to go amiss, i.e., to become what Kahneman (2011) calls 'biases' or 'cognitive illusions'. Even with mistakes, DMs cling to the biases (cognitive illusions) as long as they are *ex ante* optimum, i.e., on average optimum. They would be on average optimum because of bounded rationality, where DMs adopt heuristics to economise on cognitive cost (see Khalil, 2022; Khalil and Amin, 2023).

Temerity is an optimum heuristic given that it is highly possible for the DM to face critical situations. Therefore, the DM is ready to tolerate inevitable ex post errors of the optimum heuristic. However, these errors can be excessive. Therefore, DMs set up rules to restrain temerity, to keep it within reason. Such rules include a limit on consumption every instance of consumption.

However, DMs can do even better. They can improve wellbeing by adopting exceptions (holiday licenses) to the rule, i.e., allowing overconsumption for a specified period.

The model

A generalised principal-agent framework

This paper has so far defended three core theses, culminating in the temerity-dependent rules hypothesis. This defence is not axiom-based. This paper will make suggestions in this direction by proposing a framework, while leaving future research for a full-fledged axiom-based model of the proposed hypothesis.

The proposed framework promises to keep track of the expected cost-and-benefit of risks in a principal-agent model. In this model, the likelihood of success must be less than 100%. Otherwise, the model is useless. However, we need to modify the model considering the temerity-dependent rule hypothesis. In the modified model, the principal and the agent are the same person, not even separated by a time period. That is, we do not even have two selves, each living in a different period, à la the multiple-self approach (e.g., Fudenberg and Levine, 2006; Thaler and Shefrin, 1981).

In the proposed generalised model, the agent is the DM, while the principal is the judge that knows the true probabilities. The principal (judge) resides within the person's breast, resembling Adam Smith's impartial spectator (Smith, 1976; Khalil, 2010, 2017c). The principal (judge) keeps track of the prior. In contrast, the DM has a temerarious attitude: the DM tends to exaggerate the likelihood of success. If the DM becomes aware, as a result of experience, of how to update properly the principal's

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

prior, then the DM becomes a sophisticate. In this case, there would be absolutely no difference between the principal and the agent. We would have the standard rational DM that inhabits textbook economics.

This modified model is more general than the standard principal-agent model in an additional way. The standard model is limited in one important sense. It does not consider that the agent, when hired, must feel obligated to perform the different tasks demanded by duty, i.e., what she has promised the principal to deliver. As stated at the outset (Khalil, 2020), for standard economists it is non-conventional, if not strange, to conceive the interaction in a prisoner's dilemma setting as constrained by duty toward pertinent others – analytically similar to duty toward the future self. It is proposed here that the rational choice of players in the prisoner's dilemma game is cooperation – at least consistent with the standard economist's proposition that the rational choice of the present self in intertemporal allocation is cooperation (see also Khalil, 2026a).

This conception, however, should not be regarded as strange. Standard economists propose that the present self must feel obligated to exercise, save, eat healthy, and so on, i.e., do what it sees as the best interest of the future self while also taking care of the best interest of the present self. They should, to remain consistent, propose that players in prisoner's dilemma games must also feel obligated to keep the ex ante promises made to the other players.

If we suppose that the present self is motivated by duty toward the interest of the future self, why not extend such biological unity to social unity, i.e., equally suppose that the employee/agent is obligated to the employer/principal? Why should a thinker impose a metaphysical assumption, namely, that the biological unit (present and future selves) trumps the social unit (employee and employer selves)? Put differently, if a thinker is ready to suppose some kind of obligation between their present and future selves, why refuse to extend the same kind of obligation between employee and employer selves?

In the proposed generalised model, the way the DM can be an employee deciding whether to shirk on the job or not is analytically the same as the DM can be a present self deciding whether to cheat the future self or not. This isomorphic nature of the decision – i.e., isomorphic across two contemporaneous people or across two selves belonging to the same biological person – has many applications that cannot be discussed here (see Khalil, 2015, 2017a).

With the proposed modified model, the 'agent' and the 'principal' are the same person, where the agent is the present self while the principal is the future self. If the DM/agent (i.e., present self) decides to shirk, the principal (i.e., future self) would suffer. The standard principal-agent model is a special case, as it imposes the additional restriction that the two are separate persons. This restriction permits it to escape the issue of obligation, namely, to suppose that an employee/DM would not feel obligated toward the employer/principal, while supposing that the present self feels such obligation toward the future self. With this restriction, the standard model supposes that cheating impacts the principal negatively but only impacts the cheating DM negatively if he gets caught.

The proposed generalised principal-agent model, where the principal and the agent are the same person, allows us to clearly comprehend the choice of 'overeating' or 'undersaving'. The party that can be potentially harmed is the adjacent future self: such a self would potentially feel some discomfort by the suboptimal decision of the current self. That is, the generalised principal-agent model is nothing but individual decision-making, whereas the DM may take the gamble and enjoy an extra piece of cake or less consumption, given the likelihood that the future adjacent self would suffer.

In the proposed generalised principal-agent framework, the decision to eat or to save is *analytically* the same as the decision to shirk on the job. Both are gambles that the DM is ready to undertake, given the likelihood of success. In fact, the adjacent self in the decision of eating or saving becomes the principal in the standard model if such an adjacent self is conceived totally as separate. Then, the current self would only have to worry about the punishment, given the likelihood of success. This worry

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

is the same irrespective of whether the probable punishment is meted out by an employer or by the future self via a special mechanism.[7]

From the generalised principal-agent framework to the power of taxation

The proposed principal-agent framework can be generalised in a different way. The proposed theory – namely, DMs set up rules that limit temerity regarding over-consumption and other intertemporal allocations – is easily generalisable to constitutional rules regarding public finance. One of the major concerns of public finance is the power of taxation and its possible predatory practice by the king or sultan – as alluded to above regarding Ibn Khaldûn's methodology (1967, vol. 2: 103–111).

The concern with public finance differs from the issue of constitutional rules regarding civil liberties discussed at the outset. Regarding public finance, the issue is – e.g., in light of North and Weingast's (1989) study of the English Glorious Revolution in 1689 – who should have the power to tax: the executive (i.e., the king) or the parliament? With the change of the constitution as a result of the Glorious Revolution, the parliament was able to wrest the power to tax from the hands of the king. As North and Weingast show, the new constitutional rule ushered in fruitful institutions that encouraged investments, given that investors do not have to fear the predatory actions of the executive.

However, why is the executive more prone to predatory taxation than the parliament? That is, why do kings, more often than their respective parliaments, fall victim to excessive temerity? Given that kings cannot be easily removed, they are more prone to raise taxes under some excessive temerarious beliefs, such as that extra taxation would not lower investments and, as a corollary, undermine long-term prosperity. Given it is much easier to remove the parliamentarians by one method or another, they are sensitive to public opinion. Hence, they are less susceptible to excessive temerarious beliefs and would more likely than the King raise taxes on the basis of evidence.

In short, it is insufficient to erect rules that assist the DM to fight off excessive temerity. What matters is *credible* rules. For the constitution to specify that only the parliament can raise taxes, whereas the parliament is based on representation, the power to tax is less easily swayed by excessive temerarious beliefs than would be the case if the power rested in the hands of a hard-to-remove king.

The assumptions

The proposed generalised principal-agent model employs the following reasonable assumptions:

1. The objective functions of the DM and the principal are made up of identical time-consistent preferences;
2. The DM and the principal make decisions at discrete time τ ($\tau = t_1, \ldots, t_n$), running from the present (t_1) to retirement (t_n), where they start with the correct prior at t_1;
3. The DM and the principal differ regarding Bayesian updating of the likelihood of success starting at t_2: while the principal employs Bayesian updating, the DM forms biased priors (P^b) in the direction of temerity;
4. The DM, not the principal, is the doer, and hence, the acts are based on P^b starting t_2;
5. The DM is either unaware or, more likely, dismissive of the prognosis of the principal regarding success;
6. The DM and the principal have perfect and common information about their property rights, i.e., they concur regarding the costs and benefits of shirking or taking gambles;
7. The DM does not suffer from any loss of integrity, utility, moral disutility, blame, etc., for taking an action that may inflict pain on the principal–whereas any pain is simply the cost of the act;

[7]As the punishment mechanism detailed elsewhere (Khalil, 2020), the future self can punish the current self by withholding loans awarded to the current self. The loan mechanism is not an imaginary tool – it rather duplicates exactly how the principal or partner in the prisoners' dilemma game enforces the choice of the cooperation strategy.

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

8. The DM and the principal are risk-neutral;
9. Regarding the standard principal-agent model, if the DM is caught shirking, the principal would punish the DM, but the DM cannot retaliate.
10. The DM does not gain any skill or extra ability from learning or investing in technology on how to evade the principal's punishment because the punishment is unmitigated function of whether the principal suffers from the DM's gamble.

Decisions and Dynamics

The Principal's Command

The principal commands that the DM undertakes a binary act A_τ to take a risky consumption, a confrontation, or a gamble (A_τ^{g}) or not (A_τ^{ng}) to maximise expected benefit (Q_τ),

$$Q_\tau = P_\tau B_\tau - (1 - P_\tau)C_\tau \tag{1}$$

P_τ is the probability of success based on facts available at time τ, i.e., the probability of avoidance of detection and punishment or, in the generalised model, the avoidance of self-harm. B_τ is the benefit of the gamble, whether the gain is a good, leisure, a bribe, an embezzlement, and so on. If the principal is caught, she would have to surrender B_τ or its equivalent. In addition, the principal would have to pay C_τ, the cost of punishment in terms of imprisonment, the loss of bonus, the loss of reputation, or, in the generalised model, bodily harm and discomfort.

We may complicate this simple decision by adding, in the standard case, that the agent may have some positive altruistic preference toward the principal. Also, in the standard case, the agent may invest in technology to avoid detection/punishment.

To keep the decision simple, the principal commands

$$A_\tau^{g}: P_\tau B_\tau > (1 - P_\tau)C_\tau$$

$$A_\tau^{ng}: P_\tau B_\tau \leq (1 - P_\tau)C_\tau \tag{2}$$

For example, in the generalised model, if the expected discomfort from eating an extra serving of food is less than the expected benefit, the principal commands the extra serving. In contrast, in the standard model, the principal commands shirking if the DM's expected additional leisure exceeds the expected cost, say, of losing bonus pay.

It might be hard to imagine how the principal could command the agent to shirk. But the principal could raise the monitoring technology to catch the agent shirking, say, for 10 minutes. The principal finds that the cost of such technology is greater, at the margin, than the benefit of stopping the 10-minute shirking. This in effect means that the principal finds it beneficial to tolerate some shirking.

The DM's action

Irrespective of the principal's command, the DM (agent) reaches the decision about whether to gamble given the biased prior (P^b), according to the temerity heuristic assumption. Let us assume that a gamble was taken in a period preceding the one under focus. In this case, Pb either surpasses or fails to surpass P (the objective probability held by the principal):

$$P^b{}_\tau IA_{\tau-1}{}^{g} \begin{cases} > P_\tau IB_{\tau-1} > 0 \\ = P_\tau IC_{\tau-1} > 0 \end{cases} \tag{3}$$
$$\text{for all } \tau\ (\tau = t_2, \ldots, t_n)$$

where 'I' is the operator 'given'. That is, the DM updates the prior in an asymmetrical fashion. She updates it upward if the gamble succeeds in the previous period, but do not symmetrically update downward otherwise. That is, the heuristic engenders a (temerity) bias.

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

With this temerity heuristic assumption, the asymmetrical response would generate, on average, over time a belief that excessively exaggerates the likelihood of success.

Put differently, the temerarious DM updates upward above the principal's probability (P) upon learning of the accidental success in the previous period. However, *ceteris paribus*, the DM does not update downward below the principal's P upon learning of the accidental failure in the previous period. The DM would be a non-biased estimator, i.e., free of temerity, if her downward update fully offsets her upward update.

In this rendition, the DM reaches the principal's P in the case of failure. But the assumption need not be this strong. It is enough to stipulate that the DM may update downward below the principal's P, but not to an extent that fully offsets the upward swing in the case of success.

With enough rounds, the discrepancy between the DM's P^b and the principal's P will widen, where the DM's P^b will become higher than the principal's in an excessive manner. There can be two different mechanisms that engender the discrepancy.

First, it could be the case that a series of gambling acts that succeed each other decreases the probability of success in avoiding, in the case of eating, stomach discomfort or, in the case of shirking, getting detected and punished. While the principal properly updates the probability downward, the DM maintains the old probability. So, the DM fails to abstain from the gambling acts, considering the rising probability of getting caught or enduring self-harm.

The discussion above has already suggested the second mechanism. On the one hand, success leads the DM to exaggerate the updated likelihood of success above what is warranted by the facts. On the other hand, the DM does not sufficiently lower the likelihood of success in the face of failure. Here the DM is rather selective in the assessment of the facts to buttress excessively positive beliefs regarding the success of gambles.

Dynamics

Let us examine a stylised dynamic to demonstrate the consequence of the discrepancy between the principal's P and the DM's P^b. Table 2 shows the decisions of whether to gamble in six time periods ($\tau = 1, \ldots, 6$). The first row specifies the time period; the second row lists the principal's expected benefit based on P; and the third row registers the DM's expected benefit based on P^b.

When the expected benefit (Q_τ) is strictly positive for either self, the optimal decision is to gamble. The superscript indicates the probability upon which the decision is based. In this stylised example, the assessment of Q_τ is the same in the first period. While the assessment of Q_τ may diverge in the second period, the decision is the same: to gamble. From the third period onwards, the discrepancy of the likelihood engenders different decisions. But the action expresses the DM's decision, which is to gamble when the principal commands the contrary. In this stylised example, we assume that the DM is lucky and succeeds, which emboldens him or her further to exaggerate the likelihood of success and, correspondingly, to gamble.

If the dynamics is left alone, optimum temerity escalates to an excessive, dangerous level. Thus, the rational DM places rules—e.g., setting up a maximum level of consumption per period—to tame excessive temerity.

The thesis of the paper is accomplished. The root of rules amounts to the principal's decision to annul intertemporal dynamics of biased probability formation. This biased probability formation provides the insight that the rule, under a specified condition, can be suspended without the specter of inviting intertemporal dynamics of biased probability formation. The condition is allowing over-consumption only within specified holiday periods. The holiday license condition effectively neutralizes the intertemporal dynamics of biased probability formation. The success of overconsumption during a holiday license does not entice the DM to upward update the probability of success. Once the DM places the self within the confines of an institution, i.e., a holiday, that explicitly allows over-consumption, any positive outcome does not lead the DM to wrongly upgrade the likelihood of success.

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

Table 2. How actions become reckless

	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$	$\tau = 6$
Principal's command	$Q_1 > 0$	$Q^P{}_2 > 0$	$Q^P{}_3 \leq 0$	$Q^P{}_4 \leq 0$	$Q^P{}_5 \leq 0$	$Q^P{}_6 \leq 0$
DM's action	$Q_1 > 0$	$Q^{bP}{}_2 > 0$	$Q^{bP}{}_3 > 0$	$Q^{bP}{}_4 > 0$	$Q^{bP}{}_5 > 0$	$Q^{bP}{}_6 > 0$

In contrast, if we trace the rule to primitive present-bias preferences—which is the basis of diverse received theories of temptations—we are unable to explain why rational DMs invent holiday licenses. We would be free from the Holiday License Paradox once the analysis starts with rules, and trace them to a temerity heuristic—what this paper maintains.

Conclusion

This paper advances a basic idea. If we want to understand temptations, we should not start with tempting, present-bias preferences. Rather, we should start with rules (core thesis I). To understand rules, we should start with temerity (core thesis II). Finally, to understand temerity, we should start with optimal heuristics needed in critical decisions (core thesis III).

According to core thesis III, DMs adopt efficient temerity-as-heuristics to ensure survival in critical life-and-death situations. But such heuristics can go amiss, causing the lowering of wellbeing or fitness. Therefore, DMs can do better by erecting rules to stave off the excesses of temerity, i.e., applying rules that annul excessive temerity-based decisions in situations that are not life-and-death.

Once we have rules adopted to correct for excessive temerity-based choices (core thesis II), we can understand tempting, present-bias preferences (core thesis I). Such preferences are suboptimal only in relation to the adopted rules.

Still, what is the payoff of rejecting the received literature that conceives temptations as prior to rules? One payoff is the solution of the Holiday Licence Paradox. Insofar as received theories conceive tempting preferences as the entry point, they basically must conceive any deviation from rules as suboptimal. DMs have to basically lead a puritan lifestyle. Such a lifestyle has no room for holiday licences at the entry point of analysis.

In contrast, if we conceive temerity-as-heuristics as the entry point – where rules function as corrective technologies of excessive temerity-as-heuristics – we can easily explain holiday licences. DMs can see how such licences improve wellbeing, i.e., how the violations of rules can be beneficial as long as the violations are confined to holidays; they do not pose any threat to the rules once the holiday is over. After all, the 'enemy' is not the violation of the rules but rather intertemporal dynamics of biased probability formation.

For future research, experiments can be conducted to test the proposed hypothesis. Namely, rules are a function of temerity. It is difficult to measure the variation of rules, i.e., puritan lifestyle, across participants in the experiment. We can instead measure the variation of reckless choices (succumbing to temptations) across participants. The experiment design consists of three stages. In the first stage, the participants are tested regarding their propensity to be temerarious. The experimenter can use one of the tests developed by psychologists, e.g., Kruger and Dunning (1999). The second stage of the experiment is to screen out participants who are risk-loving. (Those participants can be paid to do other unrelated tasks to the end of the experiment). In the third stage, the experimenter offers different investment strategies with different returns and risks. While the returns on the risky options are higher than on less risky options, it should be obvious that the difference should not justify the investment in the risky options. Experimenters are given a chance to learn how to calculate 'expected value'. They are tested to make sure they realise that the expected value of the risky options is lower than the less risky options. That is, the less risky options dominate the risky ones for the participants (who are risk averse or risk neutral). The hypothesis is that the participants with high degrees of temerity will invest more

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

frequently in the risky options than those with low degrees of risk. Thus, participants with high degrees of temerity will end up with lower financial portfolios than their counterparts.

There can be many variations of this experiment design, including measuring whether temerity is correlated with risk-loving. The first task, however, is to question the standard entry point that posits rules as a function of temptations. The task is to start conceiving temptations as a function of rules.

Data Availability Statement. This paper does not include any data.

Acknowledgements. Earlier drafts benefited from the comments of Jonathan Baron, Harold Demsetz, Eric Schliesser, Leonidas Montes, James Buchanan, Yew-Kwang Ng, Ian McDonald, Haiou Zhou, and especially the comments of Tim Rakow and Steven Gardner. The paper benefited from the support of the Konrad Lorenz Institute for Evolution and Cognition Research (Altenberg, Austria) and from the Faculty of Business and Economics at Monash University. The most recent draft benefited from the comments of five anonymous reviewers. Michael Dunstan has provided valuable research assistance, and Maks Sipowicz provided useful editorial help. The usual caveat applies.

Conflict(s) of Interest. This paper is free from conflict of interest of any kind.

Funding statement. The research behind this paper has not received any funding. Open Access funding provided by the Qatar National Library

References

Ainslie G. (2001). *Breakdown of Will*. Cambridge: Cambridge University Press.
Andreou C. (2023). *Choosing Well: The Good, the Bad, and the Trivial*. Oxford: Oxford University Press.
Bakhtin M. (1965). *Rabelais and His World*. Cambridge, MA: MIT Press.
Barber B.M. and Odean T. (2001). Boys will be boys: gender, overconfidence, and common stock investment. *Quarterly Journal of Economics* **116**, 261–292.
Bhaskar V. and Thomas C. (2019). The culture of overconfidence. *American Economic Review: Insights* **1**, 95–110.
Brown P.C., Roediger H.L. and McDaniel M.A. (2014). *Make It Stick: The Science of Successful Learning*. Cambridge, MA: Belknap Press.
Brunnermeier M.K. and Parker J.A. (2005). Optimal expectations. *American Economic Review* **95**, 1092–1118.
Buchanan J. (1975). The Samaritan's dilemma. In Edmund S. Phelps (eds), *Altruism, Morality, and Economic Theory*. New York: Russell Sage, pp. 71–85.
Buehler R., Griffin D. and Ross M. (1994). Exploring the planning fallacy: why people underestimate their task completion times. *Journal of Personality and Social Psychology* **67**, 366–381.
Bénabou R. and Tirole J. (2004). Willpower and personal rules. *Journal of Political Economy* **112**, 848–887.
Camerer C.F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
Camerer C.F. and Lovallo D. (1999). Overconfidence and excess entry: an experimental approach. *American Economic Review* **89**, 306–318.
Camerer C.F., Loewenstein G. and Rabin M. (eds.) (2004). *Advances in Behavioral Economics*. New York: Russell Sage Foundation; Princeton, NJ: Princeton University Press.
Chen S., Fu R., Wedge L. and Zou Z. (2019). Non-hyperbolic discounting and dynamic preference reversal. *Theory and Decision* **86**, 283–302.
Churchland P.S. and Sejnowski T.J. (1992). *The Computational Brain*. Cambridge, MA: MIT Press.
Damasio A.R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.
Dawkins R. (1989). *The Selfish Gene*, 2nd Edn. Oxford: Oxford University Press.
Dennett D.C. (1995). *Darwin's Dangerous Idea: Evolution and the Meaning of Life*. New York: Simon and Schuster.
Dennett D.C. (2006). *Breaking the Spell: Religion as a Natural Phenomenon*. New York: Penguin.
Dewey J. (1927). Imperialism is easy. *The New Republic* **50**, 133–134.
Dhami S. (2017). *The Foundations of Behavioral Economic Analysis*. Oxford: Oxford University Press.
Dhami S. (2025). *Principles of Behavioral Economics: Microeconomics and Human Behavior*. Cambridge: Cambridge University Press.
Dietler M. and Hayden B. (eds.) (2001). *Feasts: Archaeological and Ethnographic Perspectives on Food, Politics, and Power*. Washington, DC: Smithsonian Institution Press.
Dunning D., Meyerowitz J.A. and Holzberg A.D. (1989). Ambiguity and self-evaluation: the role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology* **57**, 1082–1090.
Elster J. (ed.) (2000). *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge: Cambridge University Press.
Evans J. and Over D. (1996). *Rationality and Reasoning*. New York: Psychology Press.

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

Falassi A. (ed.) (1987). *Time Out of Time: Essays on the Festival*. Albuquerque: University of New Mexico Press.
Fischer S. (1980). Dynamic inconsistency, and the benevolent dissembling government. *Journal of Economic Dynamics and Control* **2**, 93–107.
Fitouchi L., André J.B., and Baumard N. (2023). Moral disciplining: the cognitive and evolutionary foundations of puritanical morality. *Behavioral and Brain Sciences* **46**, e293. https://doi.org/10.1017/S0140525X22002047
Fudenberg D. and Levine D.K. (2006). A dual-self model of impulse control. *American Economic Review* **96**, 1449–1476.
Galton F. (1889). *Natural Inheritance*. London: Macmillan.
Gazzaniga M.S. (ed.) (2000). *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press.
Goody J. (1982). *Cooking, Cuisine and Class: A Study in Comparative Sociology*. Cambridge: Cambridge University Press.
Gould S.J. and Lewontin R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society B: Biological Sciences* **205**, 581–598.
Grayot J.D. (2020). Dual process theories in behavioral economics and neuroeconomics: a critical review. *Review of Philosophy and Psychology* **11**, 105–136.
Gul F. and Pesendorfer W. (2001). Temptation and self-control. *Econometrica* **69**, 1403–1435.
Gul F. and Pesendorfer W. (2004a). Self-control and the theory of consumption. *Econometrica* **72**, 110–158.
Gul F. and Pesendorfer W. (2004b). Self-control, revealed preference and consumption choice. *Review of Economic Dynamics* **7**, 243–264.
Heather N. and S. Gabriel (eds.) (2017). *Addiction and Choice: Rethinking the Relationship*. Oxford: Oxford University Press.
Heffernan M. (2011). *Willful Blindness: Why We Ignore the Obvious at Our Peril*. New York: Walker.
Hodgson G.M. (2015). *Conceptualizing Capitalism: Institutions, Evolution, Future*. Chicago: University of Chicago Press.
Hoelzl E. and Rustichini A. (2005). Overconfidence: do you put your money on it? *Economic Journal* **115**, 305–318.
Khaldûn Ibn (1967). *The Muqaddimah: An Introduction to History*, **3** vols, 2nd Edn, trans. from the Arabic by Franz Rosenthal. Bollingen Series XLIII. Princeton, NJ: Princeton University Press.
Kadous K., Tayler W.B., Thayer J.M., and Young D. (2014). Individual characteristics and the disposition effect: the opposing effects of confidence and self-regard. *Journal of Behavioral Finance* **15**, 235–250.
Kahneman D. (2011). *Thinking, Fast and Slow*. New York: Farrar Straus & Giroux.
Khalil E.L. (1989). Adam Smith and Albert Einstein: the aesthetic principle of truth. *History of Economics Society Bulletin* **11**, 222–237. (Renamed: Journal of the History of Economic Thought).
Khalil E.L. (1999). Sentimental fools: a critique of Amartya Sen's notion of commitment. *Journal of Economic Behavior and Organization* **40**, 373–386. https://doi.org/10.1016/S0167-2681(99)00060-8
Khalil E.L. (2007a). Comment (Ibn Khaldûn: neither 'Muslim' nor 'Arab'). *Contemporary Sociology* **36**, 2.
Khalil E.L. (2007b). Ibn Khaldûn on property rights, The Muqaddimah: an introduction to history. *Journal of Institutional Economics* **3**, 227–238. https://doi.org/10.1017/S1744137407000677
Khalil E.L. (2009). Self-deceit and self-serving bias: Adam Smith on 'general rules. *Journal of Institutional Economics* **5**, 251–258. (It includes a long excerpt from Adam Smith's The Theory of Moral Sentiments, 1976, 156–161.).
Khalil E.L. (2010). Adam Smith's concept of self-command as a solution to dynamic inconsistency and the commitment problem. *Economic Inquiry* **48**, 177–191.
Khalil E.L. (2012). Why Europe? A critique of institutionalist and culturalist economics. *Journal of Economic Surveys* **26**, 351–372. https://doi.org/10.1111/j.1467-6419.2010.00654.x
Khalil E.L. (2013). Lock-in institutions and efficiency. *Journal of Economic Behavior and Organization* **88**, 27–36. https://doi.org/10.1016/j.jebo.2011.10.017
Khalil E.L. (2015). Temptations as impulsivity: how far are regret and the Allais paradox from shoplifting? *Economic Modelling* **51**, 551–559.
Khalil E.L. (2016). Self-deception as a weightless mask. *Facta Universitatis, Series, Philosophy, Sociology, Psychology, and History* **15**, 1–11. https://casopisi.junis.ni.ac.rs/index.php/FUPhilSocPsyHist/article/view/1377/1287
Khalil E.L. (2017a). Weakness of will and stiffness of will: how far are shirking, slackening, favoritism, spoiling of children, and pornography from obsessive-compulsive behavior? In Altman Morris (ed), *Handbook of Behavioral Economics and Smart Decision-Making: Rational Decision-Making Within the Bounds of Reason*. Cheltenham, UK: Edward Elgar, pp. 492–514. https://doi.org/10.4337/9781782549598.00041
Khalil E.L. (2017b). Making sense of self-deception: distinguishing self-deception from delusion, moral licensing, cognitive dissonance and other self-distortions. *Philosophy* **92**, 539–563. https://doi.org/10.1017/S003181911700033X
Khalil E.L. (2017c). Socialized view of man vs. rational choice theory: what does Smith's sympathy have to say? *Journal of Economic Behavior and Organization* **143**, 223–240.
Khalil E.L. (2020). The isomorphism hypothesis: the prisoners' dilemma as intertemporal allocation, and vice versa. *Journal of Economic Behavior & Organization* **76**, 735–746. https://doi.org/10.1016/j.jebo.2018.10.020
Khalil E.L. (2022). The information inelasticity of habits: Kahneman's bounded rationality or Simon's procedural rationality? *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science* **200**, 343, online publication. https://doi.org/10.1007/s11229-022-03815-7

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

Khalil E.L. (2023). Can rational choice explain hope and patience? Frustration and bitterness in The Book of Job. *International Journal of Philosophy and Theology* **84**, 55–76, https://doi.org/10.1080/21692327.2023.2172064
Khalil E.L. and Amin A. (2023). The parallelism hypothesis of cognitive economy and physiological economy: a rationality-based dual process theory. *Current Psychology* **42**, 24148–24162. https://doi.org/10.1007/s12144-022-03554-5
Khalil E.L. and Vasudevan S. (2025). Is self-deception rational? A paper submitted for publication.
Khalil E.L. (2025a). Anxiety and rationality: Allais paradox, procrastination, Keynesian expectations, and other anxiety-based deviations from rationality. *Humanities and Social Sciences Communications* **12**, 1214. https://rdcu.be/eyDXl. https://doi.org/10.1057/s41599-025-05552-x
Khalil E.L. (2025b). Real and apparent deviations from rational choice. A paper submitted for publication.
Khalil E.L. (2025c). Dual process theory: is rationality indispensable? A paper submitted for publication.
Khalil E.L. (2026a). Four theories of cooperation: the astounding failure to explain the restart effect. *Review of Behavioral Economics* forthcoming.
Khalil E.L. (2026b). To confront or not to confront? a theory of temerity. A paper submitted for publication.
Khalil E.L. (2026c). The holiday license paradox. A paper submitted for publication.
Kivetz R. and Keinan A. (2006). Repenting hyperopia: an analysis of self-control regrets. *Journal of Consumer Research* **33**, 273–282.
Kruger J. and 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**, 1121–1134.
Kőszegi B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association* **4**, 673–707.
Laibson D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* **112**, 443–477.
Luban D.J. (1999). Contrived ignorance. *Georgetown Law Journal* **87**, 957–980.
MacDonald T.K. and Ross M. (1999). Assessing the accuracy of predictions about dating relationships: how and why do lovers' predictions differ from those made by observers? *Personality and Social Psychology Bulletin* **25**, 1417–1429.
Madden G.J. and Bickel W.K. (eds.) (2010). *Impulsivity: The Behavioral and Neurological Science of Discounting*. Washington, DC: American Psychological Association.
Maio G.R., Haddock G. and Verplanken B. (2018). *The Psychology of Attitudes and Attitude Change*, 3rd Edn. Los Angeles: Sage.
Miller J.B. and Sanjurjo A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica* **86**, 2019–2047.
Moore D.A. and Schatz D. (2017). The three faces of overconfidence. *Social and Personality Psychology Compass* **11**, e12331. https://doi.org/10.1111/spc3.12331
Mulligan C.B. (1999). Galton versus the human capital approach to inheritance. *Journal of Political Economy* **107**, S184–S224.
North D.C. and Weingast B.R. (1989). Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England. *Journal of Economic History* **49**, 803–832. https://doi.org/10.1017/S0022050700009451
North D.C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge University Press.
North D.C (1991). Institutions. *Journal of Economic Perspectives* **5**, 97–112.
Olson M. (2000). *Power and Prosperity: Outgrowing Communist and Capitalist Dictatorships*. New York: Basic Books.
O'Donoghue T. and Rabin M. (1999). Doing it now or later. *American Economic Review* **89**, 103–124.
Schelling T.C. (1978). Ergonomics, or the art of self-management. *American Economic Review, Papers and Proceedings* **68**, 290–294.
Schelling T.C. (1984). Self-command in practice, in policy, and in a theory of rational choice. *American Economic Review, Papers and Proceedings*, **74**, 1–11.
Schwartz E.L. (ed.) (1990). *Computational Neuroscience*. Cambridge, MA: MIT Press.
Sharot T. (2011). *The Optimism Bias: A Tour of the Irrationally Positive Brain*. New York: Pantheon Books.
Smith A. (1976). *The Theory of Moral Sentiments*, Raphael D.D. and Macfie A.L. (eds.). Oxford: Clarendon Press.
Stanley T.D. (1991). Let's get serious about caprice. *Journal of Socio-Economics* **20**, 37–56.
Strotz R.H. (1955). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies* **23**, 165–180.
Sugden R. (2018). *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford: Oxford University Press.
Swann W.B. and Gill M.J. (1997). Confidence and accuracy in person perception: do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology* **73**, 747–757.
Talmont-Kaminski K. (2014). *Religion as Magical Ideology: How the Supernatural Reflects Rationality*. London: Routledge.
Tetlock P.E. (2005). *Expert Political Judgment: How Good Is It? How Can we Know? Princeton*. NJ: Princeton University Press.
Teubner G. and Beckers A. (2013). Expanding constitutionalism. *Indiana Journal of Global Legal Studies* **20**, 523–550. https://doi.org/10.2979/indjglolegstu.20.2.523
Thaler R.H. and Shefrin H.M. (1981). An economic theory of self-control. *Journal of Political Economy* **89**, 392–406.
Thaler R.H. and Sunstein C.R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
Thaler R.H. (2016). *Misbehaving*. New York: W.W. Norton.

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press

Wagner G.P. (2014). *Homology, Genes, and Evolutionary Innovation*. Princeton: Princeton University Press.
Waluchow W. and Kyritsis D. (2023). Constitutionalism. The Stanford Encyclopedia of Philosophy, eds. Edward N. Zalta & Uri Nodelman. https://plato.stanford.edu/archives/sum2023/entries/constitutionalism (accessed 21 January 2026).
Williamson O.E. (1985). *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press/Macmillan.

Cite this article: Khalil EL (2026). Rules and temptations: which came first? *Journal of Institutional Economics* **22**, e9, 1–19. https://doi.org/10.1017/S1744137426100411

https://doi.org/10.1017/S1744137426100411 Published online by Cambridge University Press