Annual Review of Economics

Cognitive Limitations: Failures of Contingent Thinking

Muriel Niederle[1,2,3] and Emanuel Vespa[4]

[1]Department of Economics, Stanford University, Stanford, California, USA; email: niederle@stanford.edu

[2]National Bureau of Economic Research, Cambridge, Massachusetts, USA

[3]Stanford Institute for Economic Policy Research, Stanford, California, USA

[4]Department of Economics, University of California, San Diego, California, USA; email: evespa@ucsd.edu

Annu. Rev. Econ. 2023. 15:307–28

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-091622-124733

JEL codes: C91, D91, C72, D80

Keywords

behavioral theory, experiments, cognitive economics, cognitive limitations, sure-thing principle

Abstract

In recent years, experiments have documented a new mechanism that leads to failures of profit maximization: the failure of contingent thinking (FCT). This article summarizes key experimental findings, clarifies what constitutes an FCT, and outlines how FCTs can be tested in other environments. Subsequently, we relate FCTs to recent theoretical work on cognitive limitations in behavioral economics. Finally, we connect FCTs to suboptimal behavior documented in applied environments.

1. INTRODUCTION

In recent years, experiments have documented a new mechanism that leads to failures of profit maximization: the failure of contingent thinking (FCT). This survey has two goals. First, we summarize the experimental findings. This is the purpose of the next section, which presents three examples to describe the cognitive limitations we aim to classify as an FCT. Roughly speaking, we argue that there is an FCT when an agent does optimize in a presentation of the problem that helps them focus on all relevant contingencies (i.e., contingencies in which choices can result in different consequences) but does not optimize if the problem is presented without such aids (i.e., the standard presentation). In other words, the FCT mechanism behind suboptimal behavior is that the only limitation agents suffer from is that, on their own, they sometimes focus on nonrelevant states or fail to focus on relevant ones. In our next section, we discuss the differences and similarities between the three illustrative problems and connect them to other recent examples from the experimental literature.

Our second goal is to describe how experimental findings on FCTs relate to recent theoretical work in behavioral economics on difficulties and limitations in cognitive reasoning. Even though popular behavioral models were not written focusing on FCTs, we illustrate that in some cases models can capture aspects of the difficulties of contingent thinking. We also discuss ways in which these models fall short of rationalizing FCTs.

While the experimental evidence on FCTs largely comes from the laboratory, we conclude the article by discussing how FCTs can help account for behavior in two applied environments. The first setting is a college admissions problem studied by Rees-Jones et al. (2022). That paper is inspired by data from college admissions in the United Kingdom, where suboptimal choices lead to applicants being unmatched and not attending college in the subsequent year. The second environment concerns health insurance choices. The literature (e.g., Bhargava et al. 2017) has documented a large proportion of agents who select dominated plans. Our final section connects both of these cases to FCTs and discusses why understanding the mechanism behind these mistakes is relevant for policy. The last section also provides a discussion and conclusion, with some guidance for open questions.

2. EXPERIMENTAL RESULTS

To describe evidence of FCTs, we first define what constitutes a contingency and the failure to think about it. We describe here simple cases with finite state and action sets, though in some examples these sets will be infinite.

In environments with uncertainty, elements of the state space capture relevant resolutions of uncertainty. In strategic settings, states result from considering all possible moves from others as well. Specifically, in the simplest version, let Ω be the state space, A be the action set of the agent, and $u(a, \omega_i)$ the agent's utility when taking action a in state $\omega_i \in \Omega$. State ω_i has a probability p_i of being the realized state, with $\sum p_i = 1$ and $p_i \in [0, 1]$ for $i = 1, \ldots, n$. These probabilities can be objective or subjective. A contingency C is any subset of Ω.

In environments without uncertainty, we think of p_i as the (relative) weight of state ω_i in the agent's utility function. As before, a contingency C is any subset of Ω.

For example, ω_i could be the value of a company, low or high, for $i = 1, 2$, respectively. Suppose the agent takes an action a. In the case of uncertainty, we can imagine a single company of either a low or a high value. In this case, p_1 and p_2 may represent the probability of a single company being of low or high value, respectively, with the expected utility of the agent captured by $\sum p_i u(a, \omega_i)$. In the case of no uncertainty, we can imagine many companies existing at once. In this case, p_1 and p_2 represent the shares of the total number of companies that have a low and a high value, respectively. An agent who takes an action a receives a certain (average) payoff of $\sum p_i u(a, \omega_i)$ per company. Basically, the second case is the distributional version of the first probabilistic case. We revisit these two views of states in the second example.

2.1. Failure of Contingent Thinking

We describe three examples of FCTs and discuss how we can distinguish an FCT from other cognitive mistakes. We conclude this section with a discussion on generalizing the ideas from the examples to other economic environments. The three examples consist of cases in which an FCT consists of the agent roughly not focusing on the relevant state, focusing on too few states, and focusing on too many states.

2.1.1. Example 1: Committee Voting. Consider an agent in the following voting problem. There is a jar with 7 red and 3 blue balls. One ball is randomly selected, and while the agent is aware of the jar composition, they do not know the color of the selected ball. The agent is part of a committee with three members, and the task of the committee is to select a color, red or blue, by majority vote. If the majority vote matches the ball selected from the jar, the agent receives a prize; otherwise, the agent receives nothing. The agent knows that the other two committee members are computers and knows how they are programmed to vote. One computer always votes red. The other computer votes sincerely, namely, this computer's vote matches the color of the selected ball.

Optimal behavior requires integrating information about the strategies of the computers and the rules of the game. Specifically, the agent should ask themselves under which conditions their vote matters. Given that the committee's choice is determined by majority vote, the agent is pivotal only when computers vote for different colors. The agent can then infer the color of the ball selected from the jar in that pivotal case by integrating information from the computers' strategies. Since one computer votes resolutely for red and one sincerely, this only happens when the selected ball is blue. Hence, even if there is only a 30% chance for the selected ball to be blue, voting blue is the dominant strategy.[1]

In a laboratory experiment, Esponda & Vespa (2021) find that only about 15% of participants vote optimally. A slight variation of the voting problem, studied by Esponda & Vespa (2014), shows that even after many repetitions and feedback, most participants do not vote optimally. In fact, behavior is largely consistent with naïve voting: Individuals vote for the color with more balls in the jar, ignoring information contained by the computers' rules.

There are many possible reasons for suboptimal behavior in the voting problem. For the failure of contingent thinking to be an explanation, it needs to be the case that participants are able to understand that (*a*) they should vote blue when they are pivotal, and (*b*) their vote does not matter when they are not pivotal (when both computers vote red). In a slight variation of the game, Esponda & Vespa (2014) have a sequential version in which participants learn how computers voted before they vote themselves. About three-quarters of the participants vote optimally when they know that they are pivotal. In addition, almost all participants indicate indifference when both computers voted for the same color. Hence, when participants are placed in either of the

[1]The state-space in this problem captures the possible combinations of (*a*) the color of the selected ball (*red*, *blue*) and (*b*) the possible votes of the computers [(*red*, *red*), (*red*, *blue*), (*blue*, *red*), (*blue*, *blue*)]. States in which the first (red-voting) computer votes blue have probability zero given that computer's strategy. There are two states in which the participant's vote is pivotal. Because the second computer votes sincerely, among states where the voter is pivotal, the state where the selected ball is red has zero probability, and only the state where the selected ball is blue has positive weight.

two contingencies, they are able to compute the best response; that is, they are able to make the correct inference from the computers' strategies.

Subsequently, participants are once more asked to vote simultaneously with the computers (i.e., not knowing how computers voted). Even after substantial experience with the sequential problem, only 20% of them vote optimally. This fraction is comparable to the 15% proportion of participants who vote correctly in the simultaneous case without having had prior experience with the sequential environment. Therefore, a large fraction of participants are able to vote optimally when placed in the relevant contingency but struggle when they have to determine the correct action not knowing the contingency. We characterize all those participants as suffering from a failure of contingent thinking.

The manipulation that identifies participants suffering from FCTs relies on feedback by placing participants in all relevant contingencies. An alternative, which identifies FCTs without relying on feedback, is the following change of frame, employed by Esponda & Vespa (2021). Participants choose their vote as in the simultaneous version of the game described above. However, the way their vote is used is described differently to them. Specifically, participants are explicitly told that their vote determines the majority choice only when the two computers vote for different colors. If computers vote for the same color, that color is implemented. Basically, the frame emphasizes that the agent's vote is taken into account only when the agent is pivotal. In this contingent manipulation there is no resolution of uncertainty, and the uncertainty is the same as in the original problem. The frequency of optimal voting increases from 15% to 54% percent. Hence, about 40% of participants are aided by the frame that helps them focus on the relevant contingency, even if they are not directly placed in it.

Note that, in principle, there are many reasons a participant may vote suboptimally on the original problem. In fact, in the contingent manipulation described in the previous paragraph, 46% of participants still respond incorrectly. This means that cognitive limitations beyond FCTs also play a role in understanding suboptimal choices in the original problem. However, for participants who answer correctly once they are presented with the contingent manipulation, the challenge is centered on thinking through relevant contingencies, which is why we refer to this group as suffering from an FCT. The FCT we identify in this problem is one where individuals fail to identify the correct or important contingency, namely, the state in which they are pivotal when selecting an action.

In the voting environment we just described, the optimal choice is (weakly) dominant. In the next problem we present, there is no such dominance.

2.1.2. Example 2: Acquiring-a-Company problem.

The Acquiring-a-Company problem, introduced by Samuelson & Bazerman (1985), is a variation of the classic lemons problem by Akerlof (1970). There is a company for sale that can be of high ($v_{\mathrm{H}} = 120$) or low ($v_{\mathrm{L}} = 20$) value with equal chance. There is one buyer, who can offer either a high ($p_{\mathrm{H}} = 120$) or low ($p_{\mathrm{L}} = 20$) price. When acquiring a company of value v, the buyer has profits of $1.5v$, because, say, the buyer is a better manager. The seller is a computer that knows the value of the company and is programmed to sell whenever the buyer's price is at least as large as the value.

The buyer's optimal behavior needs to account for the correlation between their offered price and the seller's action. Bidding p_{H} can lead to a large loss ($30 - 120 = -90$) if the company is of low value or to a gain ($180 - 120 = 60$) if it is of high value. Meanwhile, offering p_{L} leads to no trade if the company is of high value and to a small profit ($30 - 20 = 10$) if the company is of low value. Although there is no dominance, rational behavior (for a risk-neutral or mildly risk-loving agent) is to bid p_{L}.

Charness & Levin (2009) were the first to conduct an experiment that replaces the seller with a computer following the strategy described above, turning this game into a decision problem.[2] Their paper documents that a large proportion of participants submit $p = v_H$ even if risk elicitation indicates that they are risk neutral or risk averse.

While there are many reasons behavior is suboptimal, Martínez-Marquina et al. (2019) implement an experimental design that shows that a significant share of mistakes can be attributed to FCT. Their paper first replicates the qualitative results of Charness & Levin (2009), with approximately 20% of participants submitting the optimal prices given their risk preferences. After submitting their prices, participants are incentivized to give advice to future participants. A pattern that emerges is that participants who bid p_H tend to focus on the potential gain of 60 if the company is of high value but largely ignore the potential loss of 90 if the company is of low value. This suggests that when participants submit a price, they do not think through the consequences of their choice for both relevant contingencies, the low- and high-value company.

To identify the share of participants who make mistakes due to FCT, Martínez-Marquina et al. (2019) propose a treatment where both contingencies, the low- and high-value company, are realized. Specifically, both companies exist, and the buyer submits one bid, which is offered to each company separately. Depending on the bid, the buyer acquires only the low-value company or both the high- and the low-value company.[3,4] While in the standard version of the problem submitting a price of p_L was optimal as long as an individual was not very risk seeking, it is now a dominant strategy. To realize this, participants have to think of both companies when submitting a price, as in the standard version of the problem. In particular, they have to realize that offering price p_H results in a loss of 90 from the low-value company and a gain of 60 from the high-value company.[5] The rate of participants making an optimal choice in this modified problem more than doubles. Furthermore, most participants write advice that mentions the consequences of their choice for both companies rather than honing in on only one, as in the standard version of the problem.

In the Acquiring-a-Company problem, the optimal decision requires participants to think of all (both) hypothetical states, namely that the company can be of low or of high value, and to compute payoffs for each action-state pair. The evidence suggests that most participants instead hone in on one state only. One driver could be a cognitive limitation, which consists of individuals having an inherent problem of holding more than one state in their mind. However, many of those same participants are able to focus on two states when states are realized rather than hypothetical. We hypothesize that participants find it easier to focus on realized rather than hypothetical states. In the standard problem, it is possible to compute a possible payoff for a given action (e.g., 60 when bidding p_H if the company is of high value) without computing payoffs for that action in other states. In the problem with realized states, if one wants to know a possible payoff of bidding p_H, it is necessary to compute the payoff of this action for both states, since both states take place with certainty.

The fact that participants can do better in the problem with realized states confirms that the problem for many participants is not due to the complexity of having to handle two companies.

---

[2]The environment described by Charness & Levin (2009) is richer because the buyer is allowed to submit any positive price, not just $p \in \{v_L, v_H\}$. However, they show that after repeated play and feedback, the vast majority of prices are either p_L or p_H.

[3]An alternative way to interpret this treatment is that instead of having probabilities, the problem is presented in terms of frequencies.

[4]For ease of presentation, we describe the environment as if participants can only submit p_L or p_H. In the actual experiment, however, participants could submit other prices, which is why it is also possible to end up with none of the companies.

[5]To eliminate changes in incentives, the expected payoffs in both treatments are equalized.

Rather, the problem lies in considering hypothetical states. Since the manipulation with realized states increases the number of participants who behave as if they were thinking through all relevant contingencies, we think of this as an FCT.[6]

2.1.3. Example 3: Ellsberg problem's first question. In this example we highlight that an FCT can be present even in simple situations in which agents do not have to construct any payoff mapping from actions and states—that is, in situations in which payoffs for each action-state pair are already provided to participants. Specifically, consider the first question in the simple example from Ellsberg (1961). There is a jar with 90 balls, 30 of which are red. The other 60 are either yellow or blue, but the exact composition is not known to the agent. The agent chooses between two options whose payments depend on a randomly drawn ball. The first option pays \$10 if the ball is yellow or blue and \$0 otherwise. The second option pays \$10 if the ball is red or blue and \$0 otherwise.

Before we continue, we present the sure-thing principle (STP) due to Savage (1972). STP applies meaningfully to cases in which the set of states Ω can be partitioned in two sets A and A^c. Set A contains all states where the payoffs depend on a player's action. In contrast, states in A^c yield a payoff that is independent of the action chosen by the agent. An agent satisfies STP if their choice for the problem with states Ω is identical to their choice when the states are restricted to states A, that is, if they ignore the set A^c. While in many problems, such as the Acquiring-a-Company problem, A^c is empty, this is not the case in the Committee Voting problem. There, A is the state where the two computers vote differently and the agent is pivotal, while A^c is the set of states where both computers submit the same vote. Savage (1972, p. 21) introduces STP claiming "I know of no other extralogical principle governing decisions that finds such ready acceptance."

In the Ellsberg problem, according to STP, set $A^c = \{blue\}$, which pays the same regardless of the chosen option. Being indifferent among options in set A^c, STP claims that the agent makes the choice by focusing on set $A = \{yellow, red\}$. For an early inquiry into STP, readers are referred to Tversky & Shafir (1992). Esponda & Vespa (2021) test STP by presenting problems in two frames. For the Ellsberg problem, the standard frame corresponds to the description in the previous paragraph. In the relevant contingencies frame, participants are told that if the ball is blue, they will receive \$10. Otherwise, the payment is determined by their choice. The first option pays \$10 if the ball is yellow (and \$0 if it is red), while the second pays \$10 if the ball is red (and \$0 if it is yellow). Esponda & Vespa (2021) find that 42% of participants fail STP.[7]

If participants, as posited by Savage, want to satisfy the STP, a failure of STP is indicative of an FCT. Consider such a participant who wants to follow STP and whose "true" choice is captured in the frame that focuses on $A = \{yellow, red\}$.[8] If this participant makes a different choice when the problem is presented without partitioning states in A and A^c, they suffer from an FCT. If they would realize that both choice problems are the same, they would select the same option, but they have a problem focusing on A when presented with $\Omega = A \cup A^c$. As of now, there is no direct test on whether participants do want to follow STP [though see Esponda & Vespa (2021) for suggestive

[6]Clearly, there are other reasons beyond FCTs that might explain mistakes in the standard setting. Even though the rate of optimal choices doubles in the problem without uncertainty, there is still a large fraction of participants selecting suboptimally. These participants may have difficulties that go beyond FCTs.

[7]Notice that failure of STP is independent of failure of separability (Savage's P2 axiom), which requires a second question with different payments in A^c for testing.

[8]Note that the fact that choices are different does not, per se, indicate which choice is closer to an agent's fundamental preference. It is, however, often assumed that choices in simpler frames are closer to the agent's true choice. As $A \subsetneq \Omega$, the choice focusing on A can be considered to be the one less likely to nudge participants toward a mistake.

evidence and Nielsen & Rehbeck (2022) for how to test whether participants want to adhere to such general principles].

2.1.4. Similarities and differences between the three examples.

In the Committee Voting problem and in the Acquiring-a-Company problem the payoffs for each state and possible action are not directly provided. However, in both problems, there is evidence that many participants are capable of computing payoffs. As we have showed, in manipulations that help participants think about relevant contingencies many participants can and do compute payoffs. However, the choices of participants in the standard problem (i.e., without any help) do not reflect that participants go through these computations on their own. For instance, we suspect that if provided with the option of receiving a positive payoff for sure or a lottery that either pays this positive payoff or pays nothing, laboratory participants would prefer the former. This is essentially the choice they face in the Committee Voting problem.[9] In the Acquiring-a-Company problem, we know that if participants construct the lotteries they face, they do not end up submitting the high price (see Martínez-Marquina et al. 2019). It therefore seems that the challenges for participants lie in difficulties of thinking through the full problem on their own.

This particular challenge of having to think through the full problem without prompts is not present in the Ellsberg problem. In fact, STP is particularly useful to highlight that even in cases when payoffs are directly provided there can be FCTs. Meanwhile, the Acquiring-a-Company problem shows that FCTs can arise even when STP has no bite (because all states are payoff relevant).

More broadly, in all three problems, the manipulations that alleviate mistakes confirm that the underlying issue is an FCT. These manipulations suggest that the exact way in which agents fail at contingent reasoning differs across situations. However, all manipulations share at their core a way to focus attention on relevant states, in the extreme case by making them certain.

The three problems share a common thread: An agent who suffers from an FCT fails to profit-maximize without any help but can optimize in manipulations of the problem that help them focus on all relevant contingencies, that is, on the set A where choices can result in different consequences. In other words, the FCT mechanism behind suboptimal behavior in the original problem is that the agents, on their own, do not focus only on all relevant states.

It is relatively straightforward to see how failures of the sure-thing principle can generalize beyond the specific Ellsberg problem (see, e.g., Esponda & Vespa 2021). We next discuss how the other two examples we presented relate to other environments that have been studied in the laboratory.

2.1.5. Generalizing the Committee Voting problem.

In the voting problem, participants are not able to focus on the only relevant contingency—namely, when they are pivotal.[10] The treatment that alleviates their problem is one that highlights this state—either because the game is changed to a sequential game or because it is emphasized that the participant's vote will only be processed when their vote is pivotal and influences the outcome. The fact that this manipulation is successful confirms that the difficulties (to the extent that they are eradicated by the manipulation) stem from an FCT. This is because participants can integrate all relevant information when prompted to do so. The difficulty appears to be that without help, it simply does not occur to many

[9] If the participant votes blue, the participant receives the prize if the ball is blue (when their pivotal vote is crucial) and also when the ball is red (both computers vote red in this case). If the participant votes red, they do receive the prize if the ball is red but fail to get the prize when the ball is blue (when their vote is pivotal).

[10] Note that this contingency is potentially complex in that participants have to construct it, as it is a combination of the state of the world and the actions of others.

participants to focus on the case in which their vote is pivotal. The fact that many participants vote naïvely—that is, vote according to the more likely color of the ball—suggests that they focus on the possible color of the selected ball rather than the states containing both moves of nature as well as the moves of other players.

The Committee Voting problem shows not only that people act as if they did not take computers' votes into account but also that they are able to do so when prompted. This rules out the following alternative reasons for naïve voting in the standard problem: (*a*) Participants are not able to compute the correct action once they observe the computers' votes; (*b*) participants do take the computers' votes into account, but believe that computers' votes are erratic and do not carry information about the underlying color of the ball; and (*c*) participants want to get it right and vote for the color that matches the color of the selected ball.[11]

There are many other examples in which individuals do not seem to reason through the problem on their own. However, only a few studies have examined whether individuals are able to do so once prompted to think of relevant states.

For example, Fragiadakis et al. (2016) show that in two-player guessing games participants are (*a*) basically always able to best respond to a given strategy of an opponent and (*b*) often able to replicate their own past action in a specific game. However, individuals are much less likely to best respond to their own past action when they are not prompted to first replicate it. Such "strategic myopia" of ignoring opponents seems a good description of the behavior in the Committee Voting game as well.

The problem of strategic myopia points to another manipulation that has proven to help participants condition on the strategies of others. Several experiments have participants first predict the action of their opponent, that is, provide their beliefs about the opponent's action, before submitting their own action. Croson (1999), for instance, shows that participants who are asked to predict what the other player will do are much less likely to cooperate in a prisoners' dilemma. This suggests that in the standard version of the prisoners' dilemma game, individuals do not condition their actions on their opponent's action. However, when participants are prompted to do so, they are able to take their opponent into account (or at least more so than before).[12]

The most extreme way of alleviating strategic myopia is when the game is played in a more sequential rather than simultaneous way. In many games that we presented, individuals who suffer only from FCT are predicted to behave optimally; indeed, as we discussed, participants in many experiments are then more likely to make the correct choice. Several recent experimental papers have used analogous interventions to provide evidence of FCTs in a variety of settings. These include, among others, the work of Moser (2019) in auction-related environments, of Louis (2015) and Ali et al. (2021) in voting environments, of Calford & Cason (2021) in a public goods setting, and of Ngangoué & Weizsäcker (2021) in a market setting. In addition, Martin & Munoz-Rodriguez (2021) show that helping participants focus on relevant contingencies helps them behave optimally in a BDM (Becker–Degroot–Marschak) elicitation.

2.1.6. Generalizing the Acquiring-a-Company problem. In the Acquiring-a-Company problem, participants appear to hone in on one particular state, specifically, the state in which the company is of high value, at the expense of considering other states. This happens despite

[11] These explanations may well describe participants who vote incorrectly even in a presentation that helps with contingent reasoning.

[12] The literature has documented many instances where eliciting beliefs significantly alters play (e.g., Erev et al. 1993, Croson 2000, Rutström & Wilcox 2009, Gächter & Renner 2010). Other studies fail to reject the null hypothesis that play is not affected by eliciting beliefs (e.g., Nyarko & Schotter 2002, Costa-Gomes & Weizsäcker 2008).

their awareness that the company may be of low value and their ability to compute the payoffs in each state. The successful manipulation helps participants focus on all relevant states by making them realized rather than hypothetical. While this changes the game, it is akin to individuals playing the distribution rather than just one possible realization. Brocas & Carrillo (2022) also use this manipulation successfully, confirming that removing uncertainty helps participants focus on all relevant contingencies.

2.2. Discussion

The examples we discussed so far involve substantial simplifications of the economic environments that motivate them. In the Committee Voting problem, the participant's task is to infer information from the actions of other committee members. In the Acquiring-a-Company problem, the participants need to consider multiple states before selecting an action. One can think of this problem also as one in which participants have to be able to understand selection in that they are able to acquire all the companies worth at most as much as the bid they submitted. The Ellsberg problem consists of individuals eliminating states that do not affect outcomes—if they indeed want to make choices consistent with the sure-thing principle.

Many economic experiments, some in more complex settings, have shown that participants do not necessarily behave according to the Bayes-Nash equilibrium or make decisions that do not maximize their payoffs in ways that indicate they made mistakes. To mention a few examples, some of which we return to in the next section, this includes overbidding in common value auctions (CVAs) (see, e.g., Kagel & Levin 2002), problems of selection (see, e.g., Araujo et al. 2021), and problems of suboptimal behavior in matching environments (see, e.g., Li 2017). In many of those environments, experiments have largely focused on showing nonoptimal or nonequilibrium play. In general, they do not include treatments that allow us to precisely determine the reason behind the optimization failure. Many of these experiments are classics, and it may be time to revisit them with a more modern experimental approach that not only aims to provide clean evidence of nonequilibrium play but also is more careful in analyzing the structural or theoretical reasons behind the observed phenomenon of reasoning mistakes.

At this point we can only speculate that FCTs, which are present in simple environments, also play a role in more complex ones.[13] Meanwhile, the theoretical literature has perhaps jumped ahead of the experimental one and produced several models that rationalize classic experimental findings. In the next section we describe selected theoretical models of reasoning mistakes, explain how they relate to aspects of FCT, and discuss how their predictions deviate from documented examples of FCT.

3. BEHAVIORAL MODELS OF REASONING MISTAKES

We discuss popular behavioral models that account for some aspects of reasoning mistakes. For each model, we discuss whether, in addition to the heuristics or biases it was meant to capture, it also captures FCTs or whether, concerning FCTs, it should rather be thought of as an "as if" representation.[14] While having the right model may matter less for describing behavior, it may

---

[13]This may even apply to work that relies on the sure-thing principle but sometimes attributes failure of STP to quite complex behavioral preferences rather than to a simple failure to focus on relevant states.

[14]With an "as if" representation we mean that while the model can explain suboptimality in the standard version of a problem, it cannot rationalize participants who behave optimally when aided to think about contingencies.

have a large impact when we consider policy implications. Another aspect relevant for policy implications is whether the models describe behavior that can be clearly thought of as a mistake or instead capture biases that could be viewed as a preference.[15]

3.1. Models Capturing Updating Failures

The first class of models we consider are those that capture that economic agents do not seem to update their beliefs about states of the world based on information that is revealed during play. We first consider cases where this information is revealed by the strategies of other players and then cases in which the information is revealed by a draw from nature in decision problems.

3.1.1. Belief-based models: level-*k* and cursed equilibrium. A first set of models concerns games with several players. These models are written to capture phenomena and experimental findings that can best be described as agents failing to capture information that is revealed, in equilibrium, by the actions of others. We describe these models using the canonical setting of a CVA.

To describe the problem, as well as the essence of the level-k and cursed equilibrium models, we describe a simple CVA, used by Ivanov et al. (2010). Consider the case of two bidders, each of whom privately observes a signal x_i, for $i = 1, 2$, that is random uniformly drawn from $\{0, 1, \ldots, 10\}$. Let $x_{max} = max(x_1, x_2)$ be the highest of the two signals. Given x_1 and x_2, the ex-post common value of the item to the bidders is x_{max}. Bidders bid in a sealed-bid second-price auction where the highest bidder wins, earns the common value, x_{max}, and pays the second highest bid. In case of a tie, each bidder gets the object with equal probability.

Given a signal x_i, a player bidding b underbids, bids their signal, overbids, or bids above 10 if $b < x_i$, $b = x_i$, $x_i < b \leq 10$, or $b > 10$, respectively. Ivanov et al. (2010) show that $b(x_i) = x_i$ is the unique bid function remaining after two rounds of iterated deletion of weakly dominated bid functions.

In particular, all bid functions $b_i(.)$ with $b_i(x_i) > 10$ or $b_i(x_i) < x_i$ are removed in the first round. Underbidding is weakly dominated since the object is always worth at least x_i. Given that no one underbids, $b_i(x_i) > x_i$ is weakly dominated for any x_i, because, in case the highest bid of the other player is between x_i and $b_i(x_i)$, i makes zero or negative profits. Note that the result holds when x_i for $i = 1, 2$ is random uniformly distributed over $[0, 10]$, making it easier to describe the models, although the intuition holds for the discrete environment.

In the level-k model, level-0 (L_0) players bid in some prespecified way, and level-k (L_k) players ($k = 1, 2, \ldots$) best respond to a belief that others are level-$k - 1$ (L_{k-1}) (see, e.g., Nagel 1995). For auction settings, Crawford & Iriberri (2007) consider a version of an L_0 player, who, regardless of their signal, bids uniformly over all possible bids between the minimal and maximal value of the object. An L_1 player who best responds to such an L_0 player bids $b^{L_1}(x_i) = E(x_{max}|x_i) \geq x_i$.[16]

Normally, in a symmetric Bayes-Nash equilibrium, player 1 has to condition on winning the auction and in that case has to infer that their bid, and hence their signal, is higher than player 2's. However, when player 2 is an L_0 player, L_0's bid is independent of their signal. Therefore, nothing can be inferred about the value of the item conditional on player 1 winning the auction. Therefore, it is rational for player 1 not to infer any information from the others' action. Basically, the level-k model turns what looks like a failure of player 1 into a virtue, a best response to player 2's strategy

[15] In this section we focus on static behavioral models. We briefly discuss models that allow for learning (e.g., Esponda 2008) in our last section.
[16] In the discrete case, where $x_i \in \{0, 1, \ldots, 10\}$, we have that $b^{L_1}(x_i) = (x_i^2 + x_i + 110)/22$.

(which, however, is not an equilibrium strategy). What makes this model appealing is that apart from nonequilibrium beliefs, player 1 is selecting a strategy that maximizes their payoff and does so without any mistakes.

In a χ-cursed equilibrium (Eyster & Rabin 2005), with $\chi \in [0, 1]$, players best respond to a belief that every other player j, with probability χ, chooses a bid that is type independent and distributed according to the ex-ante distribution of j's bids and, with probability $1 - \chi$, chooses a bid according to j's actual type-dependent bid function. Thus, parameter χ captures the level of "cursedness" of the players: If $\chi = 0$, we have a standard Bayes-Nash equilibrium. If $\chi = 1$, players are fully cursed and draw no inferences about the types of other players. Based on Proposition 5 of Eyster & Rabin (2005), the following bid function constitutes a symmetric χ-cursed equilibrium (CE) in this game: A bidder who receives a signal x_i places a bid $b^{\mathrm{CE}}(x_i) = (1 - \chi)x_i + \chi E(x_{max}|x_i) \geq x_i$.

Consider the case of fully cursed ($\chi = 1$) bidders. Once more, it becomes a best response for player 1 not to infer anything from player 2's bid. Hence, just like in the level-k model, in a fully cursed equilibrium, the model turns what looks like a failure of player 1 into a virtue, a best response to player 2's strategy. When players are only somewhat cursed, players do not fully ignore the information inherent in the other players' bids, although, as long as $\chi > 0$, they would not infer "as much" as they would in a Bayes-Nash equilibrium. This model is appealing in that players are in essence still selecting a strategy that maximizes their payoff, given their cursedness.

Ivanov et al. (2010) construct an experiment using the CVA environment described above and find that these models do not capture the underlying reason for overbidding in at least this specific CVA. Basically, Ivanov et al. (2010) consider two versions of the auction game. In the standard treatment, bids are elicited for all signals. A player is allowed to bid any amount between 0 and 1 million, though the bids must be rounded to two digits after the decimal point. This is essentially a discrete version of the environment described above with the intuition of the bid functions carrying through.

In a minimum bid (MinBid) treatment, each bidder i with signal x_i is allowed to bid only between x_i and 1 million. Therefore, in MinBid, bidders cannot underbid. This completely changes the strategy of an L_1 or cursed bidder. Since their opponent is weakly overbidding by design, every player's unique best response is to bid their signal.[17] The MinBid treatment requires best-response players to, once more, make inferences based on the other players' bid, because in MinBid the bidding range is adjusted based on the signal. Basically, player i, who wins the auction with a bid $b > x_i$, has to infer that the signal of the other player is at most b, that is, in $[0, b]$. Hence, in MinBid, both cursed equilibrium and level-k thinking imply that individuals should not overbid (or overbid less, in the discrete version). Ivanov et al. (2010) find no evidence for that; in fact, if anything there is more overbidding in MinBid.

A second way to address whether players have wrong beliefs about others is to provide information on others' strategies and study whether this information changes the player's actions. Note that both the level-k and the cursed equilibrium models inherently have the feature that players are able to best respond to beliefs, so they both imply that players should be able to best respond when provided with the strategy of the other players.

However, providing a player with the strategy of their opponent creates an identification problem. An individual who changes their strategy after observing the strategy of another player may do so for reasons other than because they change their beliefs about the other player's strategy.

[17]In the continuous case, the best response to overbidding is to bid the signal. In the discrete case, our bidder can overbid, as long as the bid of their opponent is not between the signal and the bid of our bidder.

For instance, a player might learn how to play after observing another player's strategy.[18] To fix beliefs about the strategy of the other player without providing any possibility to learn from that strategy, Ivanov et al. (2010) constructed the following treatment. In Phase I of the experiment, they elicited a player's bid function by having them submit a bid for various signals against random opponents. In Phase II, players were informed that they would now play against a computer whose bidding strategy was programmed to be the participant's own Phase I bid function.

Consider a player who was overbidding in Phase I due to beliefs about the bid function of their opponent. In Phase II, when the bidder best responds to their own overbidding strategy from Phase I, they should not overbid to the same extent. However, Ivanov et al. (2010) found that many bidders were overbidding to a similar extent in Phase I and Phase II. Furthermore, there were almost equally many such bidders, even when players got to see their Phase I bid function while making Phase II bids.

The evidence found by Ivanov et al. (2010) does not show many participants whose behavior is consistent with overbidding due to inaccurate beliefs about the strategy of their opponent (either because the player is cursed or because they believe their opponent to be an L_0 player). The evidence suggests that, at least in this CVA, bidders who were overbidding were largely doing so because they were unable to compute the best response to (even their own) bid function.

The results, however, leave the door open to the possibility that overbidding is driven by an FCT. A participant who has difficulties with contingent thinking cannot compute the best response, even if they know the bidding functions of the others, because they struggle to integrate the information and think through all relevant contingencies.

Consider a participant who in Phase II of Ivanov et al.'s (2010) treatment is essentially about to submit the same overbidding bid function when bidding against a computer that uses their old bid function. If the bid function is increasing in the signal (as is largely the case), then this participant wins the auction if their bid and hence their signal is higher than that of the computer. Therefore, when submitting a bid, they should condition on the event of having the highest signal. If the bidder suffers from an FCT, this may not naturally occur to them. If their main reason to keep overbidding is that failure, we could ask whether they would place a different bid if we were to tell them the following: "Your signal is higher than that of the computer. What bid do you want to submit?" We are not aware of any such direct test to assess the role of an FCT in overbidding in CVAs.[19]

While an FCT, cursedness, and level-k thinking can all deliver the phenomenon of overbidding, the mechanism leading to overbidding is quite different. A participant who has difficulties with contingent thinking is someone who thinks that there may be information in others' strategies but on their own, when faced with all contingencies in the complex environment, does not know how to extract it. In contrast, participants who do not extract information even with manipulations that help them think contingently may either be unable to do so or believe, consistently with cursedness and level-k thinking, that there is not much information to be extracted from the strategies of others.[20]

Ivanov et al. (2010) found no strong evidence of cursedness or level-k thinking. However, they did not test for FCT, and furthermore, the signal and auction environment they used is quite

[18] Player i could, for example, have an eureka moment and understand the best action once they see it, even if they couldn't find it themselves.

[19] Koch & Penczynski (2018) present a different manipulation that, in effect, lowers the need for conditional reasoning, which reduces overbidding. For a related approach, readers are referred to Moser (2019).

[20] It is possible that participants in the laboratory may have more complicated payoff functions, which could contribute to overbidding as well. Cooper & Fang (2008) study this possibility.

unusual. It may well be that in more standard common value first-price auctions, both the cursed equilibrium and the level-*k* model may capture the essence of overbidding, which leads to the winner's curse.[21] The extent to which various mechanisms for suboptimal behavior are relevant across environments is largely an open empirical question awaiting further experiments.

3.1.2. Correlation neglect. One can imagine extending fully cursed equilibrium to situations where the other player is nature. In that case, the model boils down to a player not drawing inferences from nature's move about the state of the world. This is reminiscent of correlation neglect, which captures the idea that agents treat correlated variables as uncorrelated (see Eyster & Weizsacker 2010, Enke & Zimmermann 2019). Similar to cursed equilibrium, correlation neglect is designed to explain those reasoning mistakes that manifest themselves as a lack of updating upon observing information. As such, many FCTs clearly cannot be captured by correlation neglect.[22] However, in some cases FCTs can be directly associated with neglecting a correlation. We next provide an example with policy relevance to describe correlation neglect and connections to FCTs in further detail.

Consider a college-selection problem where an agent can apply to two out of three colleges. College A pays \$10, college B pays \$5, and college C pays \$2.5. The applicant takes a test, which provides a score that is random uniformly drawn from $\{0, \ldots, 99\}$. College A's policy is to admit any applicant with a score of at least 50. College B has a lower threshold at 45, and college C accepts all students. In a laboratory experiment using this game (which is inspired by applications to colleges in the United Kingdom), Rees-Jones et al. (2022) show that approximately 50% of participants in an experiment apply to colleges A and B, that is, submit AB. At the same time, when choosing between lotteries that are the result of an application to AB and to AC respectively, participants prefer the safe lottery implied by choosing AC to the risky lottery implied by choosing AB. This suggests that applying to colleges A and B is a mistake for many participants.[23]

Participants who apply to colleges A and B despite their more conservative risk preferences make a mistake that can be captured by correlation neglect. The participant's test score decides admissions to both college A and college B, while college C admits all students. Therefore, there is a correlation between admission decisions for colleges A and B. If the participant neglects this correlation and instead treats the admissions of A and B as independent, they might prefer applying to A and B over applying to A and C.[24]

We consider three broad classes of explanations for this mistake, which result in behavior consistent with neglecting a correlation. The first is literal correlation neglect. A participant who suffers from literal correlation neglect actually believes that the admissions to colleges A and B are independent.[25] Second, more mundanely, it could be that individuals understand the correlation, that

[21]For extensive evidence on the winner's curse in common value first-price auctions, readers are referred to Kagel & Levin (2002).
[22]For example, correlation neglect cannot account for why agents focus on some states rather than all states especially when states are hypothetical (see Martínez-Marquina et al. 2019).
[23]In fact, the experiment by Rees-Jones et al. (2022) is motivated by the fact that in the United Kingdom many applicants end up not being offered admission to any of the colleges they applied to.
[24]For further details, readers may consult Rees-Jones et al. (2022), who report a larger set of treatments than we focus on here. For instance, one of their manipulations is that applicants draw a random test score for each school, keeping the distribution of admission thresholds fixed. Agents who take into account the correlation of admission decisions across schools should react as it changes. However, they document that most participants do not change their behavior.
[25]This can happen, for example, if the participant's subjective representation of the problem is one in which there is a different test for each college and test results are independent of one another.

is, do not neglect it but are simply unable to compute it.[26] While this seems implausible, it can be tested by asking participants what the possible exam scores are for someone rejected by college A.

Third, the problem could be the result of an FCT. Applying to college A as a first choice is rather obvious.[27] If the applicant's score is high enough to grant admission to college A, the choice of the second college is irrelevant. To pick their second choice, the applicant should thus focus on the contingency in which they are rejected from college A. The probability of being rejected at college A but admitted to college B is 5%. However, ignoring this crucial contingency, the student may just consider the ex-ante probability of being admitted to college B, which is 55%. Participant suffering from an FCT may clearly be aware that there is a correlation between school choices, but they do not know how to use this information because on their own they do not focus on the relevant contingency.

Basically, the mechanism behind the mistake to apply to AB rather than to AC can be, aside from problems of updating and computational mistakes, either literal correlation neglect or an FCT. However, these last two explanations lead to different policy implications. If the student suffers from literal correlation neglect, it might be useful to reinforce that both schools use the same exam. This could perhaps be achieved by having a centralized place where schools look up the exam score rather than sending scores to each school independently. If, on the other hand, the student suffers from an FCT, the policy intervention should aim to help them think through the relevant contingency. In fact, the student's ability to solve the problem in that contingency is what defines the mechanism as being purely an FCT. A possible intervention could place the participant directly in the relevant contingency, using the sequential choice framework introduced above. Instead of placing students in the contingency in which their first-choice school rejected them, they could be subjected to a contingency treatment similar to the one we discussed in the Committee Voting problem. One could tell participants that they should select a first-choice school; then, only in case their first choice does not accept them, their second-choice school will be processed and used to determine their second application.[28]

The three models of level-*k*, cursedness, and correlation neglect were geared to account for failures of updating based on information inherent in the action of other players, or of nature. This is achieved by individuals believing that others act in a non- (or less) strategic way because they are level-0 players or are cursed. Another way is ignoring the possibility of inference, either because the players themselves are cursed or because they neglect correlation. In a loose way, for these models, all game forms that are strategically equivalent (e.g., changes in framing) are likely not to make different predictions in terms of reasoning mistakes. However, we have seen that for an agent with limited contingent reasoning, it may make a huge difference if the agent is put into the contingency rather than asked to decide everything ex ante. That is, for such an agent, the game form may be very relevant.

3.2. Obvious Dominance

A strategy is obviously dominant if, for any deviating strategy, from any information set in which the strategies diverge, the best possible outcome from the deviating strategy is no better than the worst possible outcome under the considered strategy (Li 2017). Obvious dominance captures

[26] Readers are referred, for example, to Möbius et al. (2022) for failures and difficulties with updating, and to Benjamin (2019) for an overview.

[27] In fact, the proportion of participants not selecting A as the first college of the problem above in Rees-Jones et al.'s (2022) treatment is negligible.

[28] Rees-Jones et al. (2022) report on an intervention that goes in the direction of emphasizing the contingent reasoning approach and find that it is somewhat useful in affecting individuals' choices. More extremely, Bó & Hakimov (2020) find more truth telling in iterative deferred acceptance mechanisms than in standard ones.

situations in which a cognitively limited agent can recognize a dominant strategy. The cognitive limitations are, in fact, directly related to difficulties in thinking contingently. Namely, the cognitively limited agent is unable to condition on other players' actions (including nature) when comparing their strategies. They are unable to make statements like "conditional on the other player selecting action *a*, my payoff from action *b* is *x*."

Clearly, obvious dominance does not explain all FCTs, since those also occur in situations without a dominant strategy (e.g., the first question of the Ellsberg problem). However, it could be that, in cases with a dominant strategy, FCTs are captured by the lack of obviousness.

Turning a game with an obviously strategy-proof (OSP) dominant strategy into a similar one that has a dominant strategy that is not obvious involves specific changes to the game. We speculate that these changes affect the chance for an agent who suffers from an FCT to find the dominant strategy. For example, bidding one's value in a private-value second-price two-player auction is a dominant strategy, but it is not obviously dominant. The best possible outcome from a deviating bid (higher than one's value) can result in a positive payoff (if the other bidder has a low value and places a low bid), while the worst possible outcome from bidding one's value is 0 (if the other bidder places a higher bid). In contrast, in the ascending clock auction, bidding one's value is an obviously dominant strategy.[29] Following a large literature, Li (2017) provides experimental evidence supporting this prediction.[30]

While obvious dominance can account for some FCTs, it fails to account for others. Consider the Committee Voting problem we presented in the previous section. One of 7 red and 3 blue balls is randomly selected. The agent and two computers vote for red or blue, and if the majority vote coincides with the color of the selected ball, the agent receives a positive payoff. Recall that one computer always votes red and the other votes for the true color of the ball. It is an obviously dominant strategy to vote blue. The best payoff from deviation is the positive payoff, but so is the worst payoff from voting blue! Voting blue guarantees the positive payoff.[31] Yet, the majority of participants vote suboptimally.[32]

Why does obvious dominance sometimes predict behavior quite well and sometimes not? Some OSP games place agents in payoff-relevant contingencies directly. This is the case for the clock auction described above. However, other OSP games do not help agents focus on all relevant contingencies. This is the case with the voting problem. We suspect that whenever the game form is changed so as to put agents on specific contingencies, and if that makes dominant strategies obviously dominant, as in the auction example, obvious dominance correlates with optimal behavior. In cases in which the game form is changed in a manner that does not help agents focus on relevant contingencies, we suspect it is less likely to capture behavior. However, beyond the voting example given above, this remains an open experimental question.

3.3. Focusing on Specific States

One manifestation of FCTs that is relevant, for example, in the work of Martínez-Marquina et al. (2019) is that individuals do not seem to take all states into account, especially when states are

[29]If the current price is below the agent's valuation, the payoff from quitting now ($0) is no better than the payoff from quitting when the price reaches the valuation. If the current price is above the valuation, the best payoff from staying in the auction is not higher than the payoff from quitting immediately.

[30]An agent with difficulties in thinking contingently may also find the ascending auction setting easier because it helps them focus on a relevant contingency as opposed to thinking through contingencies by themselves.

[31]If the state is red, both computers vote red, making the agent's vote of blue obsolete and delivering a positive payoff. If the state is blue, the agent's vote makes blue the majority vote and hence guarantees a positive payoff.

[32]Readers are referred to Esponda & Vespa (2021) for details and to Esponda & Vespa (2014) for an example in which voting for blue is not obviously dominant.

uncertain. The next two types of models have the feature that when making decisions, agents are predicted to hone in on one or some subset of relevant states at the expense of other states. We describe how each model shares features or makes predictions that are in line with FCT. While both models have broad applications beyond FCT, we highlight results that cannot be taken into account but are specifically designed to showcase the role of FCTs.[33]

3.3.1. Salience. Salience theory describes a bottom-up attention procedure in which attributes of options act as stimuli that trigger selective recall from memory of typical values of those attributes (Bordalo et al. 2012, 2013, 2020, 2022). To illustrate salience theory as well as its connection to FCTs, we present a simple common-consequence Allais problem. We describe how salience can account for changes in behavior when a problem is changed in such a way as to correlate outcomes. We then provide a manipulation of the problem which provides a connection to the STP that salience fails to capture.

Consider an agent who is asked to choose between the following two lotteries. Lottery 1 pays \$10 with 11% chance and \$0 otherwise. Lottery 2 pays \$50 with 10% chance and \$0 otherwise. The "minimal" state space (Bordalo et al. 2012) is the set of distinct payoff combinations that occur with positive probability. For instance, lottery 1 may pay \$10 and lottery 2 may pay \$50. The minimal state space is {(\$10, \$50), (\$10, \$0), (\$0, \$50), (\$0, \$0)}. A salience function orders the states depending on underlying salience properties. A state that has a higher contrast in payoffs is more salient. Hence, the state (\$0, \$0) has minimal salience and the most salient state is (\$0, \$50). According to the salience model, the agent inflates weights attached to the most salient states and deflates those attached to the less salient ones. Inflating state (\$0, \$50), the agent may therefore be more likely to prefer the risky lottery 2.

A crucial aspect of the problem described above, which we will refer to as the lottery frame, is that the outcomes of the two lotteries are independent from one another. Alternatively, consider a problem where the outcomes of the lotteries are correlated. In this "urn" frame there is an urn with 1 red, 10 yellow, and 89 blue balls. One ball is randomly selected, but the color is not revealed to the agent. A blue ball always pays \$0. In option 1, the agent receives \$10 if the ball is red or yellow. Option 2 pays \$50 if the ball is yellow and \$0 if the ball is red. The lotteries implied by these options are the same as the lotteries in the lottery frame. However, because the payoffs depend on the draw of a ball from an urn, the lotteries are now correlated. Hence, the minimal state space excludes (\$0, \$50). Salience theory therefore predicts that individuals are more likely to select option 1 in the urn frame than lottery 1 in the lottery frame.[34]

We now describe a change in the problem that affects an agent who suffers from FCTs. The manipulation, however, does not affect an agent whose behavior is described by salience. Consider the following change in the presentation of the urn problem. Note that a blue ball offers the same payment regardless of which option is chosen. The payments of the options only differ when the selected ball is either yellow or red. In other words, this problem can be studied through the lens of STP. To do so, consider the case of an agent who faces the following contingent urn frame. The agent is told that they will get paid \$0 if the ball is blue.[35] The only task for the agent is to make

[33]There is a large set of theories that were developed to rationalize patterns of choices in problems like the Ellsberg question we presented as Example 3. Most of such theories rationalize deviations by modifying preferences. Since we focus on agents who would like to satisfy STP, in this section we describe recent behavioral theories that may rationalize choices that fail STP as a cognitive reasoning mistake. For a summary of preference-related theories, readers may consult Dhami (2016).

[34]Bordalo et al. (2012) provide suggestive evidence that is consistent with this prediction (see also Frydman & Mormann 2016).

[35]Alternatively, we assume that if the ball is blue the agent is indifferent toward getting paid \$0 from having selected option 1 or option 2.

the following contingent choice in case the selected ball is yellow or red. Option 1 pays $10 if the ball is yellow or red, while option 2 pays $50 if the ball is yellow and $0 if it is red. Note that the agent makes the choice in the contingent urn frame without knowing the color of the selected ball. Nonetheless, the contingent urn frame focuses the choice on the colors, whereby different choices have different consequences. The minimal state space is the same in the contingent urn frame and in the urn frame, so (in principle) salience theory does not predict that agents would make a different choice in these two presentations of the problem.[36]

If choices in the contingent urn frame differ from choices in the urn frame, STP is violated. Nevertheless, about 30% of participants make different choices in the urn and in the contingent urn frame (see Esponda & Vespa 2021). Those participants exhibit an FCT, assuming that all of them would like to satisfy STP. Hence, an FCT can occur even when the set of minimal states does not change and when salience theory does not predict a difference.

At the end of this section, we provide an example of an FCT that neither salience nor sparsity, which we describe next, can account for. This is despite the fact that, in principle, the example is in an environment in which both salience and sparsity could be expected to make predictions.

3.3.2. Sparsity. Sparsity envisions an agent who builds a simplified model of the world, considering only the variables of first-order importance (Gabaix 2014). Sparsity assumes that the agent is rational given the contingencies they take into account. The selection of states that receive weight is the solution of a maximization problem, so the agent is more likely to take into account relevant states and ignore nonrelevant ones. This has two implications at odds with the findings on FCTs. First, sparsity would probably predict that agents follow STP. This is because thinking about contingencies is costly, which means that the agent would likely focus only on payoff-relevant contingencies. This implies that sparsity may predict that we would not observe violations of STP as described in the previous section. Second, if under their sparse model the agent considers all payoff-relevant contingencies, there should be no difference between behavior in the deterministic and the probabilistic Acquiring-a-Company problem.

In the following paragraphs we return to a previous example in which salience and sparsity fail to account for behavior observed in experiments. Specifically, consider the voting example from Section 2.1, in which an individual who suffers from an FCT may not understand that they should vote blue when they do not know the votes of the two computers (where a vote of blue neutralizes the computer who votes red and hence renders the "honest" computer as the pivotal voter). However, such an agent realizes they should vote blue if they are put in the contingency where their vote matters (i.e., one computer voting red and one blue).

Salience, in its original model, does not include any difficulties individuals may have in constructing the payoff-relevant state space. Rather, salience is a model explaining why participants may put too much (or too little) weight on some (non)salient states. However, once payoffs have been computed, the most salient state is the one where the voter is pivotal.

Sparsity, like salience, does not include any difficulties individuals may have in constructing the payoff-relevant state space. Rather, sparsity is a model that explains why individuals focus too much on some states and ignore others. Once again, the only important state is the one where the voter is pivotal.

---

[36]We say "in principle" because salience can be augmented to capture differences between the urn frame and the contingent urn frame. Following Bordalo et al. (2022), frames may differ because they make some attributes of the problem prominent and others hidden. If one assumes that the contingent urn frame makes states in which the consequences of the choices differ (yellow and red) more prominent, this may predict a difference across the two frames.

Therefore, both models predict voters to be equally able to vote optimally in all three treatments of the Committee Voting problem, be it the original game (without any aids), the sequential treatment, or the contingent treatment.

To summarize, all three classes of models, while failing to fully account for the essence of an FCT, capture some aspects of this cognitive limitation. Salience and sparsity capture that agents who suffer from FCT sometimes focus only on a few states rather than on all payoff-relevant ones. Obvious dominance and models on updating failures capture that agents who suffer from FCT sometimes fail to construct the relevant state and fail to focus on the relevant states only.

4. CONCLUSIONS, APPLICATIONS, AND OPEN QUESTIONS

An agent who suffers from an FCT is able to maximize their profits when problems are presented in a way that focuses their attention on relevant contingencies. However, the agent does not behave optimally without such aids. This may be because they do not focus on the relevant contingency (as in Example 1, the Committee Voting problem), focus on too few contingencies (as in Example 2, the Acquiring-a-Company problem), or focus on too many contingencies (as in Example 3, the first question of the Ellsberg problem). We presented experimental evidence for all three cases, indicating that in such problems many participants fail to make optimal decisions but succeed in treatments with manipulations that help them focus on relevant contingencies.

Note that our examples also vary how direct or how simple it is to compute the relevant contingency and its associated payoffs. We provided examples of FCTs in cases where the payoffs for each contingency were directly described (as in Example 3, the first question of the Ellsberg problem), in cases where individuals have to compute the payoffs associated with each contingency (as in Example 2, the Acquiring-a-Company problem), and in the case in which individuals have to construct the relevant contingency in the first place (as in Example 1, the Committee Voting problem).

On the experimental side, many important questions remain open. First, what is the set of (at least somewhat computationally) simple problems where FCTs are the main reason for suboptimal behavior? In the Committee Voting problem, almost all participants are able to answer questions that directly help with the contingent reasoning in the problem at hand. For instance, Esponda & Vespa (2014) ask participants to respond (in an incentivized manner) whether they can change the choice of the committee if the selected ball corresponds to the color most represented in the jar. The answer is no, because in such a case both computers are programmed to vote for such color, and the participant is not pivotal. The vast majority of participants answer the question correctly. However, most of these participants fail to play optimally when they subsequently face the actual voting problem. This suggests that participants understand pieces of the problem but have difficulties putting them together. In other words, problems in which participants need to construct the (nontrivial) payoff mapping between actions and states are likely to lead to FCTs. However, there is no clear guidance on how to measure such a "complexity of construction" aspect.

In addition, little is known about how important FCTs are when mistake sources are richer than those in the simple environments we discussed. For example, Ivanov et al. (2010) showed that participants have a hard time best responding to a known strategy of an opponent in a specific second-price common-value auction. Does this imply that FCTs are largely irrelevant in more difficult problems? Clearly, when the problem is very complex (e.g., it requires solving complicated equations), FCTs, if present, may be of second order compared to cognitive problems of computation. We expect FCTs to become first order in computationally simple environments in which people behave optimally with help. These could be problems that participants can solve once they are in a specific contingency but fail to solve without help that emphasizes contingencies. Nevertheless, is it obvious that, say, the winner's curse in more standard first-price CVAs is due to cursedness, level-*k* thinking, or correlation neglect rather than being the result of an FCT?

Finally, the role of feedback is relatively unexplored. On the one hand, feedback can help by forcing participants to think through specific contingencies that took place. However, feedback is often not transparent. The quality of feedback can depend on the environment and on the quality of the agent's own choices.[37] While the experimental literature has made some progress in understanding the role of feedback, it is not well established what types of feedback would help participants overcome FCTs.[38]

From a theoretical perspective, we explored the connection between FCTs and well-established behavioral notions. Understanding the mechanisms behind suboptimal choices more precisely, and in particular the extent to which FCTs are responsible for such reasoning mistakes, can be useful to develop better theories. As we discussed, some theories can be used to rationalize the same mistake committed by an agent suffering from an FCT (even if the actual reason for the mistake is not literally captured by the theory). But from a policy perspective, there is a clear advantage in having a model that directly addresses the participants' underlying problems. There are some recent efforts in this direction. Piermont (2021), for instance, models an agent whose capacity to engage in hypothetical thinking is captured by their ability to recognize implications between hypotheses. For example, in the Committee Voting problem, this translates to an agent who does not understand the relationship between the state and the possibility of being pivotal.[39] Similarly, Cohen & Li (2022) present an extension of cursed equilibrium to sequential games. This allows for predictions where players neglect information content in hypothetical events, while making correct inferences from observed events.

Difficulties with contingent thinking can play an important role in everyday problems. Earlier we mentioned the college admission problem studied by Rees-Jones et al. (2022). Suppose that the mechanism behind suboptimal behavior is that agents act "as if" the admissions exam was independent between schools.[40] That is, they mistakenly think that admission to one school is unrelated to the chance of being admitted by other schools, while, in fact, admission to schools is guided by the same exam. In this case, the policy recommendation would be to reinforce the existing correlation. For instance, a message could be "Recall that both schools use the same exam, so schools' decisions are not independent from each other."

Alternatively, if the source behind the reasoning mistake is an FCT, the difficulty lies in thinking about the crucial contingency. That is, the admissions problem is a problem where the individual has to construct and then focus only on that one relevant contingency, namely, that their first choice school rejected them.[41] In this case, a policy to help students could be to include the following message: "The second school to which you apply is only relevant if the first school rejects

[37] Esponda (2008) introduces the notion of behavioral equilibrium, in which agents who have difficulties understanding selection in the data that they collect end up not behaving according to the Nash equilibrium in the long run.

[38] Enke (2020) and Araujo et al. (2021) offer experimental evidence pointing to participants having difficulties in dealing with selection, and Esponda & Vespa (2018) and Barron et al. (2019) discuss difficulties with endogenous selection in the feedback (i.e., feedback that depends on the quality of the agent's choices). Finally, Esponda et al. (2021) suggest that for an agent who fails at thinking contingently, feedback may help only if it makes their mistake extremely transparent.

[39] Readers may consult Piermont & Zuazo-Garin (2020) for a setting with a more general formulation.

[40] For additional examples, readers are referred to Hassidim et al. (2021) and Shorrer & Sóvágó (2022).

[41] In this sense, an FCT in the college admissions problem is similar to the FCT in the Committee Voting problem.

you. Before you think about your second choice school, put yourself in the following hypothetical situation: Your first-choice school rejected you."

Hence, the optimal policy in the college admission problem depends on the mechanism behind the mistake. Of course, it could be that different students suffer from different problems, making perhaps a message combining both points the optimal one. Our point is that more work is needed to be able to provide the best recommendation to college admission systems.

A similar situation arises in the choice of health insurance. Suppose that a health insurance plan can be characterized by a premium (a certain payment) and a deductible (the amount of health expenditures paid by the agent prior to plan coverage). To compare plans, an agent needs to think of states (possible health expenditures) and then construct how much they would end up paying for each plan. Liu & Sydnor (2022) document that health insurance plans in the United States often include dominated options. In fact, they report that in roughly half of the firms in their data set, there is a high-deductible plan that dominates all other plans. This would not be an issue if dominated plans were not selected. However, using data from a large US firm, Bhargava et al. (2017) show that the majority of employees, in fact, select a dominated plan. Their paper uses an experimental design with AmazonTurk participants to explore mechanisms. After replicating baseline suboptimal choices, they conduct a treatment in which participants are aided to compute the consequences of possible health scenarios (unhealthy, moderately healthy, very healthy). With this type of aid, the proportion of participants selecting dominated options decreases from 48% to 18%. An interpretation of these results through the lens of FCTs suggests that suboptimal choices are large when participants have to construct the payoff mapping between states and different insurance policies by themselves and have to keep all possible health states in mind.[42] Yet a manipulation that focuses participants on all three relevant health states, which in addition may help them construct the relevant payoffs, seems crucial in selecting optimal insurance plans. However, more work is needed to assess whether such a policy would indeed be helpful in many cases.

To conclude, FCTs can be an important mechanism behind suboptimal behavior not only in abstract problems but also in applications of everyday choices. However, to fully measure the extent to which this mechanism is responsible for errors, a large number of open questions need to be answered using theory and experiments as well as field applications.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We would like to thank Lukas Bolte, Russell Golman, Kirby Nielsen, Vlasta Rasocha, and Alex Rees-Jones for detailed comments on previous versions of this article, as well as Gonzalo Arrieta, Maxim Bakhtin, Tony Fan, Shengwu Li, Yucheng Liang, Alejandro Martinez-Marquina, Carl Meyer, Markus Mobius, Lea Nagel, and Sevgi Yuksel for comments that helped us organize the presentation of this survey. We also thank our coauthors who helped shape how we think of the issues discussed in this article.

[42] In this sense, the FCT in the health insurance problem is similar to the one in the Acquiring-a-Company problem.

LITERATURE CITED

Akerlof GA. 1970. The market for lemons: quality uncertainty and the market mechanism. *Q. J. Econ.* 84(3):488–500

Ali SN, Mihm M, Siga L, Tergiman C. 2021. Adverse and advantageous selection in the laboratory. *Am. Econ. Rev.* 111(7):2152–78

Araujo FA, Wang SW, Wilson AJ. 2021. The times they are a-changing: dynamic adverse selection in the laboratory. *Am. Econ. J. Microecon.* 13(4):1–22

Barron K, Huck S, Jehiel P. 2019. *Everyday econometricians: selection neglect and overoptimism when learning from others*. Discuss. Pap. SP II 2019-301, Wiss. Berl. Sozialforsch., Berlin, Ger.

Benjamin DJ. 2019. Errors in probabilistic reasoning and judgment biases. In *Handbook of Behavioral Economics: Foundations and Applications*, Vol. 2, ed. BD Bernheim, S DellaVigna, D Laibson, pp. 69–186. Amsterdam: North-Holland

Bhargava S, Loewenstein G, Sydnor J. 2017. Choose to lose: health plan choices from a menu with dominated option. *Q. J. Econ.* 132(3):1319–72

Bó I, Hakimov R. 2020. Iterative versus standard deferred acceptance: experimental evidence. *Econ. J.* 130(626):356–92

Bordalo P, Gennaioli N, Shleifer A. 2012. Salience theory of choice under risk. *Q. J. Econ.* 127(3):1243–85

Bordalo P, Gennaioli N, Shleifer A. 2013. Salience and consumer choice. *J. Political Econ.* 121(5):803–43

Bordalo P, Gennaioli N, Shleifer A. 2020. Memory, attention, and choice. *Q. J. Econ.* 135(3):1399–442

Bordalo P, Gennaioli N, Shleifer A. 2022. Salience. *Annu. Rev. Econ.* 14:521–44

Brocas I, Carrillo JD. 2022. Adverse selection and contingent reasoning in preadolescents and teenagers. *Games Econ. Behav.* 133:331–51

Calford EM, Cason TN. 2021. *Contingent reasoning and dynamic public goods provision*. ANU Work. Pap. Econ. Econom. 2021-679, Aust. Natl. Univ., Canberra

Charness G, Levin D. 2009. The origin of the winner's curse: a laboratory study. *Am. Econ. J. Microecon.* 1(1):207–36

Cohen S, Li S. 2022. Sequential cursed equilibrium. arXiv:2212.06025 [econ.TH]

Cooper DJ, Fang H. 2008. Understanding overbidding in second price auctions: an experimental study. *Econ. J.* 118(532):1572–95

Costa-Gomes MA, Weizsäcker G. 2008. Stated beliefs and play in normal-form games. *Rev. Econ. Stud.* 75(3):729–62

Crawford VP, Iriberri N. 2007. Level-*k* auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica* 75(6):1721–70

Croson RT. 1999. The disjunction effect and reason-based choice in games. *Organ. Behav. Hum. Decis. Process.* 80(2):118–33

Croson RT. 2000. Thinking like a game theorist: factors affecting the frequency of equilibrium play. *J. Econ. Behav. Organ.* 41(3):299–314

Dhami S. 2016. *The Foundations of Behavioral Economic Analysis*. Oxford, UK: Oxford Univ. Press

Ellsberg D. 1961. Risk, ambiguity, and the savage axioms. *Q. J. Econ.* 75(4):643–69

Enke B. 2020. What you see is all there is. *Q. J. Econ.* 135(3):1363–98

Enke B, Zimmermann F. 2019. Correlation neglect in belief formation. *Rev. Econ. Stud.* 86(1):313–32

Erev I, Bornstein G, Wallsten TS. 1993. The negative effect of probability assessments on decision quality. *Organ. Behav. Hum. Decis. Process.* 55(1):78–94

Esponda I. 2008. Behavioral equilibrium in economies with adverse selection. *Am. Econ. Rev.* 98(4):1269–91

Esponda I, Vespa E. 2014. Hypothetical thinking and information extraction in the laboratory. *Am. Econ. J. Microecon.* 6(4):180–202

Esponda I, Vespa E. 2018. Endogenous sample selection: a laboratory study. *Quant. Econ.* 9(1):183–216

Esponda I, Vespa E. 2021. *Contingent preferences and the sure-thing principle: revisiting classic anomalies in the laboratory*. Work. Pap., Univ. Calif., Santa Barbara

Esponda I, Vespa E, Yuksel S. 2021. *Mental models and learning: the case of base-rate neglect*. Work. Pap., Univ. Calif., Santa Barbara

Eyster E, Rabin M. 2005. Cursed equilibrium. *Econometrica* 73(5):1623–72

Eyster E, Weizsacker G. 2010. *Correlation neglect in financial decision-making*. Discuss. Pap. 1104, DIW Berlin, Berlin, Ger.

Fragiadakis DE, Knoepfle DT, Niederle M. 2016. *Who is strategic?* Work. Pap., Stanford Univ., Stanford, CA

Frydman C, Mormann M. 2016. *The role of salience and attention in choice under risk: an experimental investigation*. Work. Pap., Univ. South. Calif., Los Angeles

Gabaix X. 2014. A sparsity-based model of bounded rationality. *Q. J. Econ.* 129(4):1661–710

Gächter S, Renner E. 2010. The effects of (incentivized) belief elicitation in public goods experiments. *Exp. Econ.* 13(3):364–77

Hassidim A, Romm A, Shorrer RI. 2021. The limits of incentives in economic matching procedures. *Manag. Sci.* 67(2):951–63

Ivanov A, Levin D, Niederle M. 2010. Can relaxation of beliefs rationalize the winner's curse? An experimental study. *Econometrica* 78(4):1435–52

Kagel JH, Levin D. 2002. *Common Value Auctions and the Winner's Curse*. Princeton, NJ: Princeton Univ. Press

Koch C, Penczynski SP. 2018. The winner's curse: conditional reasoning and belief formation. *J. Econ. Theory* 174:57–102

Li S. 2017. Obviously strategy-proof mechanisms. *Am. Econ. Rev.* 107(11):3257–87

Liu C, Sydnor J. 2022. Dominated options in health insurance plans. *Am. Econ. J. Econ. Policy* 14(1):277–300

Louis P. 2015. *The barrel of apples game: contingent thinking, learning from observed actions, and strategic heterogeneity*. Tech. Rep., Univ. Zurich, Zurich, Switz.

Martin D, Munoz-Rodriguez E. 2021. *Misperceiving mechanisms: imperfect perception and the failure to recognize dominant strategies*. Work. Pap., Northwest. Univ., Evanston, IL

Martínez-Marquina A, Niederle M, Vespa E. 2019. Failures in contingent reasoning: the role of uncertainty. *Am. Econ. Rev.* 109(10):3437–74

Möbius MM, Niederle M, Niehaus P, Rosenblat TS. 2022. Managing self-confidence: theory and experimental evidence. *Manag. Sci.* 68(11):7793–8514

Moser J. 2019. Hypothetical thinking and the winner's curse: an experimental investigation. *Theory Decis.* 87(1):17–56

Nagel R. 1995. Unraveling in guessing games: an experimental study. *Am. Econ. Rev.* 85(5):1313–26

Ngangoué MK, Weizsäcker G. 2021. Learning from unrealized versus realized prices. *Am. Econ. J. Microecon.* 13(2):174–201

Nielsen K, Rehbeck J. 2022. When choices are mistakes. *Am. Econ. Rev.* 112(7):2237–68

Nyarko Y, Schotter A. 2002. An experimental study of belief learning using elicited beliefs. *Econometrica* 70(3):971–1005

Piermont E. 2021. Hypothetical expected utility. arXiv:2106.15979 [econ.TH]

Piermont E, Zuazo-Garin P. 2020. Failures of contingent thinking. arXiv:2007.07703 [cs.AI]

Rees-Jones A, Shorrer RI, Tergiman CJ. 2022. *Correlation neglect in student-to-school matching*. NBER Work. Pap. 26734

Rutström EE, Wilcox NT. 2009. Stated beliefs versus inferred beliefs: a methodological inquiry and experimental test. *Games Econ. Behav.* 67(2):616–32

Samuelson WF, Bazerman MH. 1985. The winner's curse in bilateral negotiations. In *Research in Experimental Economics*, Vol. 3, ed. VL Smith, pp. 105–37. Bingley, UK: Emerald Group

Savage LJ. 1972. *The Foundations of Statistics*. New York: Dover

Shorrer RI, Sóvágó S. 2022. *Dominated choices in a strategically simple college admissions environment: the effect of admission selectivity*. Work. Pap., Pa. State Univ., University Park

Tversky A, Shafir E. 1992. The disjunction effect in choice under uncertainty. *Psychol. Sci.* 3(5):305–10