An Interventionist's Guide to Exotic Choice

Reuben Stern*

Forthcoming in *Mind*

Abstract

In this paper, I use interventionist causal models to identify some novel Newcomb problems, and subsequently use these problems to refine existing interventionist treatments of causal decision theory. The new Newcomb problems that stir trouble for existing interventionist treatments involve so-called "exotic choice" — i.e., decision-making contexts where the agent has evidence about the outcome of her choice. I argue that when choice is exotic, the interventionist can adequately capture causal-decision-theoretic reasoning by introducing a new interventionist approach to updating on exotic evidence. But I also argue that this new updating procedure is principled only if the interventionist trades in the typical interventionist conception of choice for an alternative Ramseyan conception. I end by arguing that the guide to exotic choice developed here may be useful in some everyday contexts, despite its name.

1 Introduction

Newcomb problems exemplify scenarios in which evidential decision theorists say there is reason to abstain from taking an action that causal decision theorists say

*Kansas State University – Manhattan, KS – https://sites.google.com/view/reubenstern/home – reuben.stern@gmail.com.

there is no reason not to take. More formally, when a Newcomb problem arises, x-ing dominates the agent's other option(s) according to causal decision theory, but is disfavored by evidential decision theory. Consider Big Tobacco's Dream.

> **Big Tobacco's Dream (BTD)**: Suppose that you live in a world that Big Tobacco once hoped was actual — namely, a world in which the correlation between smoking and lung cancer is entirely explained by the prominence of a genetic condition that causally promotes smoking habits and lung cancer. You would hate to get cancer, but would enjoy every other possible consequence of smoking, and are considering taking up the habit. Should you smoke?

When confronted with BTD, causal decision theorists say that you should smoke because smoking dominates abstaining. Their thought is that there is nothing that you can do now to causally influence whether you get lung cancer, and you'd prefer to smoke both if it turns out that you get lung cancer and if it turns out that you don't. Meanwhile, evidential decision theorists say that there is cancer-related reason not to smoke because smoking provides evidence that you will get lung cancer (provided that the correlation between smoking and lung cancer persists given everything that you know).[1] Their thought is that it doesn't matter that you'd prefer to smoke no matter whether you get lung cancer because smoking affects the probability that you will get lung cancer — i.e., $P(cancer|smoke) > P(cancer|\neg smoke)$.

The aim of this paper is not to settle the controversy about how agents should choose when confronted with Newcomb problems, but is rather to develop a ver-

[1]The parenthetical is required to block the evidential decision theorist's "tickle defense" of smoking.

sion of causal decision theory that captures the causal-decision-theoretic response to Newcomb problems in an elegant and simple way. The strategy is (i) to use interventionist causal models to identify some novel Newcomb problems, and (ii) to use these new Newcomb problems to refine interventionist treatments of causal decision theory. The particular Newcomb problems that stir trouble for existing interventionist treatments of causal decision theory involve so-called "exotic choice" — i.e., decision-making contexts wherein the agent has evidence about the outcome of her choice. I argue that when choice is exotic, the interventionist can adequately capture causal-decision-theoretic reasoning by introducing a new interventionist approach to updating on exotic evidence. But I also argue that that this new updating procedure is principled only if the interventionist trades in the typical interventionist conception of choice for an alternative Ramseyan conception. The Ramseyan alternative agrees with the typical interventionist conception except when choice is exotic.

2 The Interventionist Turn

I said in the last section that an aim of this paper is to develop a version of causal decision theory that captures causal-decision-theoretic reasoning in an elegant and simple way. But why not do the same for evidential decision theory? This exercise is trivial because evidential-decision-theoretic reasoning is easily captured by the claim that agents should maximize conditional expected utility (CEU) — i.e., that agents should opt for whatever option, x, maximizes expected utility when defined as follows, where $P(Y = y|X = x)$ corresponds to the conditional probability that state y will obtain given that x obtains and where $V(X = x, Y = y)$ corresponds

to the value of the outcomes associated with taking action x in state y.[2]

$$CEU(x) = \sum_{y} P(Y = y|X = x)V(X = x, Y = y)$$

Causal decision theorists cannot straightforwardly develop their decision theory in terms of CEU because there are non-causal correlations between actions and states that causal decision theorists deem irrelevant for deliberation, but that affect CEU calculations insofar as they affect what weights are used.[3] For example, even though smoking doesn't causally promote lung cancer in BTD, CEU calculations give more weight to the value of the outcomes that arise from smoking in worlds where you get lung cancer than the outcomes that arise from smoking in worlds where you do not get lung cancer because the probability of lung cancer is higher if you smoke than if you don't. As a result, if you hate cancer enough, the CEU of abstaining will be greater than the CEU of smoking, despite the fact that smoking doesn't causally promote cancer.

While there are various attempts to spell out causal-decision-theoretic reasoning, there has recently been a move towards using interventionist causal models to develop causal decision theory.[4] There are several reasons for this trend,[5] but

[2]CEU is intended to be partition-invariant in the sense that it applies no matter how one carves up the states. So while X must be partitioned such that its cells are the options on the agent's menu, Y can be partitioned however one chooses.

[3]Though causal decision theorists cannot straightforwardly develop their theories in terms of CEU, Bradley (2018) and Joyce (1999) discuss how causal decision theorists can develop partition-invariant versions of their theories that utilize distinct conceptions of supposition (not defined in terms of conditional probability).

[4]See Easwaran (forthcoming), Hitchcock (2016), Meek and Glymour (1994), Pearl (2009), and Stern (2017).

[5]See Hitchcock's (2016) defense of what he calls "causal decision metatheory" — i.e., the thesis that when engaging with a decision problem, we should use causal models to make explicit our assumptions about the causal structure of the problem, as well as the question that we are asking.

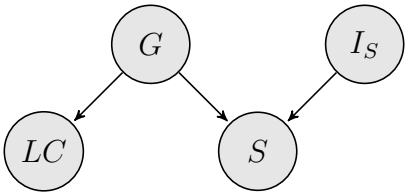
perhaps the most celebrated is that the interventionist can attain causal-decision-theoretic verdicts (e.g., that you should smoke) while preserving the evidential decision theorist's insight that you should choose whatever option is evidence for the best outcome[6] — i.e., that you should maximize conditional expected utility.[7] In order to accomplish this feat, the interventionist construes the agent's options as *interventions to act*, where the "intervention to act" is an exogenous deterministic cause of the act (rather than the act itself) that, given the axioms of the graphical approach to causal modeling, must be evidentially irrelevant to any variables that are not causally downstream from the act.[8] Thus, the following Figure 1 depiction of BTD implies that the intervention to smoke (I_S) is evidentially irrelevant to whether you have the gene (G) and whether you have lung cancer (LC), and you can thus opt to smoke without worrying about any evidential effect on G or LC when your options are treated as interventions.

Figure 1: Intervening to Smoke

[6]Meek and Glymour (1994) famously stress this point.

[7]It is worth noting that there are also ways to use interventionist causal models to develop causal decision theory such that it cannot be put in terms of maximizing CEU. For example, one can use Pearl's (2009) *do*-calculus to argue that $P(y|x)$ should be replaced with $P(y|do(x))$ when calculating expected utility, where $P(y|do(x))$ is not understood in terms of conditional probability (despite appearances) and is rather understood as the probability that y obtains given a novel kind of interventionist supposition that x obtains. Here, I adopt Meek and Glymour's (1994) approach according to which causal-decision-theoretic verdicts are attained simply by conditioning on the intervention to x rather than x itself, but every point that I make in what follows about the standard interventionist's mistreatment of exotic choice can likewise be made against the backdrop of a decision theory that utilizes the *do*-calculus. Doing so would simply require a lot of translation. See Pearl (1993) for a discussion of how one can move between the two approaches.

[8]We will see in what follows that this is not quite right when choice is exotic, even though it is right when choice is not exotic.



How does this follow from the causal modeling axioms? Allow **V** to denote the set of variables over which the relevant probability distribution and causal graph are defined.[9] The chief axiom of the causal modeling framework — namely, the Causal Markov Condition (CMC) — implies constraints on which probability distributions over **V** are compatible with which causal graphs over **V**. The CMC is a generalization of Reichenbach's (1956) Common Cause Principle that captures the sense in which causes "screen off" their effects.[10] Its implications can be neatly summarized in terms of *d-separation*.

According to the CMC, if two variables, X and Y, are *d-separated* by a (possibly empty) set of variables, Z, in some causal graph, then X and Y must be probabilistically independent of each other conditional on any assignment of values over Z in any probability distribution that is compatible with that graph. X and Y are *d-separated* by Z exactly when every path between X and Y is *blocked* by Z, where a path between X and Y is *blocked* by Z exactly when:

1. the path between X and Y contains a non-collider that is in Z, or,

[9]When I refer to a causal graph here, I am referring to a directed acyclic graph — i.e., a graph whose nodes are linked by directed edges (or arrows) in which there are no cycles.

[10]If you are unfamiliar with the language of causes "screening off" their effects, the idea is that if X and Y are not cause and effect but are nevertheless correlated because they share some common cause(s), then X and Y are not correlated given any assignment of values to their common cause(s). For example, we all take the correlation between sunscreen sales and ice cream sales to cease to persist when we condition on the weather being a particular way. In the language of "screening off," we all treat the weather as screening off sunscreen sales from ice cream sales. As we see below, this is implied by the CMC.

2. the path contains a collider, and neither the collider nor any descendant of the collider is in Z.

If you are unfamiliar with graphical causal models, then this language of "d-separation" and "colliders" is probably foreign. But since a collider is just a common effect of two variables along an undirected path[11] — e.g., S along the Figure 1 path, $LC \leftarrow G \rightarrow S \leftarrow I_S$ — the CMC can be parsed as saying that any two variables represented in some causal graph must be probabilistically independent of each other unless (i) they share a (direct or indirect) common cause, (ii) one is a (direct or indirect) cause of the other, or (iii) they are both (direct or indirect) causes of some common effect that has been conditioned on. If you understand this much about the CMC, then you understand everything about graphical causal models that is required for comprehension of this paper.

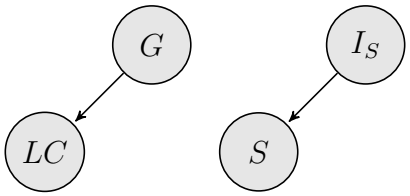
How does the CMC imply that your intervention to smoke (or not smoke) is evidentially irrelevant to whether you get lung cancer? Provided that you don't already know whether you will smoke as you decide whether to smoke (i.e., that you haven't conditioned on the collider, S), you must regard whether you intervene to smoke as probabilistically independent (and evidentially irrelevant) from whether you have the genetic condition and whether you get lung cancer.[12] Thus when you intervene to smoke, you effectively make yourself smoke in a way that does not

[11] An undirected path is just a a sequence of variables such that there is an arrow (going in either direction) between each variable and the next. Intuitively, causal arrows *collide* along paths at *colliders*.

[12] The Causal Markov Condition may not generally imply constraints on an agent's rational subjective probability function when she spreads her credence across multiple causal hypotheses. In this paper, I focus on decision-making contexts in which the agent does not spread her confidence across multiple causal hypotheses in order to abstract away from this difficulty. Thus when I speak of variables in terms of their "probability," I can be interpreted as referring to the agent's subjective credences towards these variables that are themselves informed by the objective causal probabilities that partially comprise the lone causal hypothesis under consideration. See Stern (2017) for extensive discussion of these issues.

depend on whether you have the genetic condition or whether you get lung cancer. For this reason, it is sometimes helpful to conceptualize the effect of conditioning on an intervention variable in terms of setting the intervened upon variable to a particular value in a way that (i) severs any causal dependencies between the intervened upon variable and its (non-intervention) parents, but (ii) leaves intact the dependencies between the intervened upon variable and any of its effects. This is neatly summarized when we depict the intervention as "arrow-breaking" — i.e., as breaking the dependence of S on G as follows.

Figure 2: Arrow-breaking Interventions



3 Colliding Newcomb Problems

By implying that your intervention to smoke is probabilistically independent from whether you get lung cancer, the CMC secures the causal-decision-theoretic dominance argument for smoking. That is, when your options are $intervene_{smoke}$ and $intervene_{\neg smoke}$ and the states are *cancer* and $\neg cancer$, you are guaranteed to prefer $intervene_{smoke}$ if you maximize CEU because $P(cancer|intervene_{smoke}) = P(cancer|intervene_{\neg smoke})$, thus allowing the rational choice to be determined by your relative value assignments — i.e., your preference for intervening to smoke

over intervening to abstain in both worlds where you get lung cancer and worlds where you don't get lung cancer.

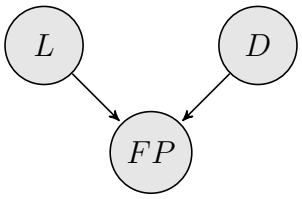
This is not the only use to which the CMC can be put in the context of developing causal decision theory. We can also use the CMC to search for new cases that drive a wedge between causal decision theorists and evidential decision theorists — e.g., new Newcomb problems. The trick to finding a Newcomb problem is to identify a case in which an agent values taking some action over the alternative(s) no matter whether some state obtains, where the relevant state is correlated with the act, but not causally downstream from the act. BTD has taught us that this can happen when the state and act are correlated because of some common cause, or when the state and act are correlated because the state is causally upstream from the act — viz., you face a Newcomb problem no matter whether you construe the states as *cancer* and $\neg$*cancer* or as *gene* and $\neg$*gene* (since your preference for worlds in which you smoke is robust across either partition and neither partition is causally downstream from whether you smoke). But the CMC also suggests some other ways that states can be correlated with an agent's action, but not causally downstream from an agent's action, all of which involve conditioning on colliders (or their descendants).[13]

In order to see how conditioning on a collider can induce a correlation, it is helpful to consider a simple case in which the phenomenon is intuitive. Consider Fiona, who eats lunch and dinner at two separate restaurants, each with its own

[13]It is important to note that the CMC does not by itself imply any probabilistic dependencies. Rather, it implies independencies, and leaves it open which dependencies obtain. Typically, the dependencies permitted by the CMC *do* obtain, but they sometimes do not. This is why the Causal Faithfulness Condition is considered a valuable heuristic in the context of using observational data to identify causal structure, where the Causal Faithfulness Condition states that the *only* independencies that obtain between variables in a graph are those implied by the CMC.

distinct food supplier. Intuitively, whether the food supply at the lunch restaurant is contaminated (L) has no evidential bearing on whether the dinner restaurant is contaminated (D). But now suppose that Fiona experiences symptoms of food poisoning (FP) the next day. If Fiona then calls up the lunch restaurant and confirms that their supply was not contaminated, it is clear that Fiona should become more confident that the dinner restaurant's supply was contaminated. Thus once Fiona has conditioned on the collider, FP, she should regard $L = \neg contaminated$ as evidence that $D = contaminated$.

Figure 3: The Food Poisoning Collider



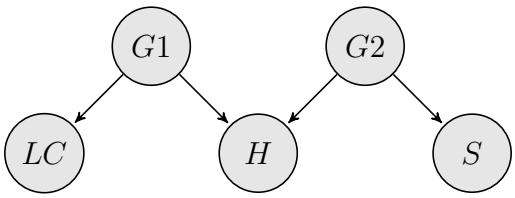
The food poisoning collider does not yet provide us with a Newcomb problem since it does not involve a choice, but it is easy to exploit this kind of reasoning to arrive at one.[14] Consider Muggsy's choice.

> **Muggsy's Choice (MC)**: On Muggsy's planet, there are two separate genetic conditions that cause shortness: one that disposes people to smoke, and another that disposes people to get lung cancer. None of these symptoms causes any other. Muggsy doesn't know whether he has either genetic condition, but he knows that he is short. Muggsy

[14]To my knowledge, this is the first explicit mention of a colliding Newcomb problem in the literature. For example, the possibility of this problem is not covered in Easwaran's (forthcoming) classification of Newcomb problems, even though Easwaran classifies Newcomb problems in terms of graphical causal models.

would find smoking unpleasant, but he really would hate to get lung cancer. Should Muggsy smoke?

Figure 4: The Causal Structure of MC



$S = \{smoke, \neg smoke\}$

$G1 = \{gene_1, \neg gene_1\}$

$G2 = \{gene_2, \neg gene_2\}$

$LC = \{cancer, \neg cancer\}$

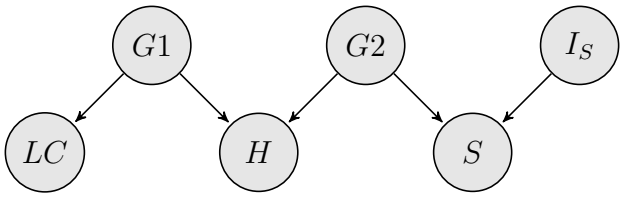
$H = \{short, \neg short\}$

Here, just as Fiona's knowledge that she was poisoned gives her reason to believe that it's more likely that dinner was poisoned in the event that lunch wasn't, Muggsy's knowledge that he is short gives him reason to believe that it's more likely that he has $gene_1$ in the event that he doesn't have $gene_2$. And since abstaining from smoking raises the probability that he doesn't have $gene_2$, it likewise raises the probability that he has $gene_1$ (relative to what he knows), and, in turn, raises the probability that he will get lung cancer. Thus Muggsy has evidential-decision-theoretic reason to smoke even though there is a clear causal-decision-theoretic dominance argument for abstaining, given Muggsy's distaste for

smoking.[15]

What do standard interventionist treatments of causal decision theory imply for MC? Exactly what they should — i.e., that there is no reason for Muggsy to smoke. The reason is that even though S is evidentially relevant to LC given that $H = short$, I_S (below) is evidentially irrelevant to LC because S is a collider on the only path between S and LC and Muggsy has *not* conditioned on any value of S. Thus if Muggsy maximizes CEU while treating his options as interventions, the causal-decision-theoretic dominance argument goes through successfully.

Figure 5: Intervening in MC



MC exemplifies a new kind of Newcomb problem that hasn't been discussed in the literature. But MC poses no obstacles for standard interventionist treatments of causal decision theory. In fact, interventionists can hang their hat on MC insofar as their theory not only manages to capture the causal-decision-theoretic response to MC, but also utilizes the causal modeling machinery that enables us to find MC in the first place. So it may seem that colliding Newcomb problems like MC only strengthen the case for the standard interventionist treatment of causal decision theory. But not so fast. In the next section, we will see that colliding Newcomb problems stir trouble even for interventionists when choice gets exotic.

[15]Remember (i) that apart from his concern about getting lung cancer, Muggsy would rather not smoke because he finds it unpleasant, and (ii) that smoking doesn't cause lung cancer.

4 Exotic Choice Problems

When an agent has evidence about the outcome of her choice, her choice is exotic. It has long been thought that such choices stir trouble for causal decision theory.[16] Consider the following exotic choice.

> **Future Medical Results 1 (FMR1)**:[17]
>
> Though you would hate to get lung cancer, you know that you would enjoy every other possible consequence of smoking, and are considering taking up the habit. An oracle tells you that smoking causes lung cancer by causing your lungs to blacken, but that the effect of smoking on lung cancer is entirely mediated by whether your lungs blacken. The oracle also gives you the results of a future medical test, revealing that your lungs will unfortunately blacken. Should you smoke?

Because there are many versions of causal decision theory, it is somewhat hard to pin down what causal-decision-theoretic reasoning favors here.[18] For example, some standard versions of causal decision theory (e.g., Sobel 1980) yield the result that there is reason to abstain because the expected utility of an action should

[16]The precise nature of this trouble is disputed. Some (e.g., Egan 2007 and Price 2012) argue that there are exotic counter-examples to standard versions of causal decision theory while others (e.g., Bales 2016, Hitchcock 2016, Rabinowicz 2009, and Sobel 1980) argue that these cases provide reason to favor some versions of causal decision theory over others. Still others (e.g., Lewis 1982) maintain that our intuitions about these cases are too foggy to be used as ammunition in the fight over which normative decision theory is right. Here, following Hitchcock (2016), my strategy (roughly speaking) is to grant Egan and Price's intuitions about what is rational, and to use interventionist tools to develop a decision theory that makes good on their intuitions.

[17]The case is called Future Medical Results 1 because there will be a sequel in what follows.

[18]FMR1 has the same structure as Lewis's (1982) "pauper's problem." See Bales (2016) and Rabinowicz (2009) for discussion of how several versions of causal decision theory apply to the pauper's problem.

be calculated in correspondence with what the chances *would* be were you (counterfactually) to take the action in question, and the chance of getting lung cancer *would* be greater were you to smoke than were you to abstain (even if it is *actually* certain). But other standard versions of causal decision theory (e.g., Lewis 1981) vindicate the dominance argument for smoking on roughly the grounds that your fate is sealed when it comes to any effect (causal or otherwise) that smoking exerts on the state of your lungs.

As far as intuitions are concerned, it seems clear (at least to this author) that you should go ahead and smoke. After all, you already know that your lungs will blacken no matter what you do. Why not savor the pleasures of the cigarette? The problem with this response, at least for traditional causal decision theorists, is that it's hard to defend on principled grounds when you're in the business of construing causal dependence (rather than evidential dependence) as that which matters for decision theory.[19] This is because whether smoking *causes* lung cancer in the external world is independent of what we know (including our knowledge that our lungs will blacken).[20]

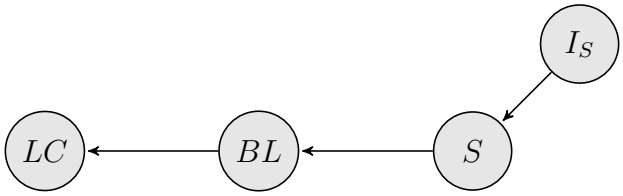
Hitchcock (2016) has noticed that this tension is not present for the interventionist causal decision theorist who maintains that agents should maximize CEU while treating their options as interventions. The reason is that the interventionist

[19]The problem here is not formal. It has already been mentioned that Lewis develops his theory so that it vindicates the dominance argument. There may likewise be ways to develop other prominent versions of causal decision theory (e.g., Joyce's 1999 version) so that it delivers the intuitive result. The problem, developed below, is rather that the causal decision theorist can't do so in a principled way under the assumption that objective causal dependence (rather than evidential dependence) is what matters for decision theory.

[20]This is obviously true when causal dependence is construed objectively. Price (2012) uses this to argue that causal dependence should be construed subjectively (i.e., as agent-relative and knowledge-dependent) if causal dependence is relevant to rational choice. Price is in the minority here — i.e., few causal decision theorists (if any) grant the subjectivity of causal dependence to Price.

does not choose between causal dependence and evidential dependence, and rather simply maintains that we should make choices in line with what the intervention to act is evidence for. Since the intervention to smoke is not evidence for lung cancer given our knowledge that our lungs will blacken, the interventionist's decision theory straightforwardly gets the intuitive result that you should smoke. To see this, consider the CMC's application to Figure 6. Since BL d-separates I_S from LC, the CMC implies that I_S is evidentially irrelevant to LC given that $BL = \textit{blackened}$. This means that there is no reason not to smoke by the interventionist's lights — i.e., if you treat your options as interventions and condition on the fact that your lungs will blacken while maximizing CEU, then you successfully take stock of the fact that your lungs will blacken no matter what you choose, and the dominance argument for smoking goes through successfully.[21]

Figure 6: Intervening in FMR1



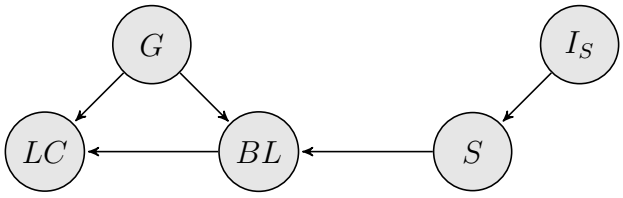
[21]Hitchcock (2016) develops this proposal in detail. He rightly emphasizes that it is important to condition on $BL = \textit{blackened}$ in the manipulated distribution — i.e., the distribution that is arrived at by conditioning on the fact that you are intervening. If you condition on $BL = \textit{blackened}$ in the unmanipulated distribution — i.e., the distribution that is arrived at by conditioning on the fact that there is no intervention — you do not take stock of the fact that your lungs will blacken no matter how you choose, and instead take stock of the fact that your lungs would blacken by default were you to abstain from making any choice at all. The dominance argument for intervening to smoke does not go through in the latter case. According to the interventionist, it is principled to condition on $BL = \textit{blackened}$ in the manipulated distribution (rather than in the unmanipulated distribution) because you intervene when you make a choice.

But the fact that standard interventionist treatments of causal decision theory sometimes deal nicely with exotic choice doesn't mean that they always do. Consider the following slight variant of FMR1, where genes and future medical results mix.

Future Medical Results 2 (FMR2):

Though you would hate to get lung cancer, you know that you would enjoy every other possible consequence of smoking, and are considering taking up the habit. An oracle tells you that smoking causes lung cancer by causing your lungs to blacken, but that the effect of smoking on lung cancer is entirely mediated by whether your lungs blacken. She also mentions that there is a genetic condition that disposes people's lungs to blacken, while (via a direct causal path) making them less susceptible to lung cancer. Finally, the oracle gives you the results of a future medical test, revealing that your lungs will blacken. Is there any reason not to smoke?

Figure 7: Intervening in FMR2



FMR2 is just like FMR1 but with one catch; there is now a genetic common cause of BL and LC. By including this common cause in FMR2, BL is transformed into a collider along a path that links I_S and LC — namely,

$LC \leftarrow G \rightarrow BL \leftarrow S \leftarrow I_S$ — and the CMC therefore allows LC and I_S to be correlated in the probability distribution that results from conditioning on $BL = \textit{blackened}$.[22] Intuitively, once you have learned that your lungs will blacken, intervening to abstain raises the probability that they blacken for other reasons (e.g., that you have the genetic condition) in the same way that Fiona's discovery that lunch wasn't poisoned raised the probability that dinner was the culprit. Since the operative genetic condition provides its carriers with a resistance to lung cancer, this means that standard interventionist treatments of causal decision theory (including Hitchcock's 2016 explicit treatment of exotic choice) depict you as having cancer-related reason to abstain from smoking.

As far as evidential-decision-theoretic reasoning is concerned, this gets things right. After all, the argument that there is cancer-related reason to abstain when confronted with FMR2 is of roughly the same form as the evidential-decision-theoretic argument that there is cancer-related reason for Muggsy to smoke when confronted with MC (insofar as both trade on non-causal correlations that arise upon conditioning on colliders). But the causal decision theorist should disagree. That is, since the correlations that arise are *not* reflective of any causal influence that S exerts on LC, causal-decision-theoretic reasoning speaks in favor of the conclusion that there is no cancer-related reason not to smoke.[23] Thus we again have a colliding Newcomb problem on our hands, only this time the causal decision theorist cannot rest content with standard interventionist treatments of causal-

[22]This difference between the two cases can be explained in terms of d-separation. Because BL blocks the only path between I_S and LC in Figure 6, the CMC mandates that I_S and LC are independent given any value of BL in FMR1. But because BL is a collider between I_S and LC along one path in Figure 7, BL does *not* block every path between I_S and LC, and the CMC therefore does *not* mandate that I_S and LC are independent given any value of BL in FMR2.

[23]Remember that the only difference between FMR2 and FMR1 is the inclusion of the genetic common cause of BL and LC. No new causal influence of S over LC was introduced.

decision-theoretic reasoning (since these treatments deliver results that do not square with causal-decision-theoretic reasoning).

What goes wrong for standard interventionist accounts of rational choice? When choice is not exotic, colliders pose no problems for interventionists since a collider must be a common effect of (and thus causally downstream from) the intervention to act in order for conditioning on the collider to induce a dependence between the intervention to act and its non-effect(s). But when choice is exotic, the agent has evidence about some variable that is causally downstream from the intervention to act (as you do in both FMR1 and FMR2), and conditioning on this evidence can induce a spurious correlation between the intervention and its non-effect(s) (as it does in FMR2). Moreover, this phenomenon is not unique to FMR2. We can imagine even simpler cases where downstream colliders stir trouble for interventionists.[24] Consider Fallen Door (FD).[25]

> **Fallen Door (FD)**: You know that a burglar plans to target your neighborhood tonight. His method is simple: he will pick at random

[24]I make this point in what follows with an example that is wholly distinct from the smoking cases discussed above, but we can also make this point with a smoking case. It's widely known that low birth-weight children born to smoking mothers have a lower infant mortality rate than low birth-weight children born to non-smoking mothers. This correlation may initially strike you as bizarre, but it is easily explained in terms of a simple collider structure — i.e., there are multiple independent causes of a child's low birth-weight, and a low birth-weight child is better off (i.e., less likely to die) in the event that the child's low birth-weight is caused by its mother's smoking habit than the other possible (more fatal) causes. (See Pearl 2016 for a more thorough explanation of this "birth weight paradox" in terms of a collider.) Now suppose that you are in the early stages of pregnancy and an oracle tells you that your child will suffer a low birth-weight. Should you take up smoking? Given that your child's birth-weight will be low, intervening to smoke decreases the probability that your child will die of the other (more fatal) causes and therefore consititutes *evidence* that your child will be (relatively) healthy. But the causal decision theorist should regard this evidential correlation as irrelevant to your choice since the other causes of low birth-weight are not causally downstream from whether you smoke.

[25]Nevin Climenhaga is the author of this example. He independently presented it on social media while this paper was under review. I am glad that this left enough time to include it in the final manuscript (with his permission). The version that appears here is truncated and paraphrased.

among the 1000 houses in your neighborhood by rolling a 1000-sided die, and then he will break down the door of that house (if need be) and will steal everything that he can.[26] Wondering whether it will be your house, you consult your crystal ball that infallibly shows you what your front porch will look like tomorrow morning. You see to your dismay that your front door has been knocked down. But you also realize that the burglar is not the only person who can knock down your door — viz., you can kick it down yourself right now if you want. You'd find the experience of kicking the door somewhat unpleasant since it'd hurt your foot a bit, but you'd really hate for your stuff to be stolen. Should you endure the mild pain and kick down the door?

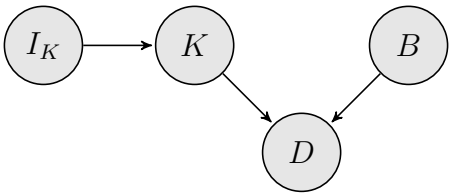
Though the probability that you will be burgled given your evidence that your door will be broken is plausibly quite high, the probability that you will be burgled given your evidence that your door will be broken *and that you intervene to kick down your own door* is quite low.[27] Thus given what you know, intervening to kick the door down strongly reduces the probability that you will be burgled. Does this mean that there is burglary-related reason for you to kick down the door? Since the correlation that arises between whether you intervene to kick and whether you're burgled does not reflect any *causal* influence that you exert over B, causal-decision-theoretic reasoning speaks in favor of refraining from kicking. That is, according to the causal decision theorist, there should be no reason at all to endure the pain of the kick when you know full well that you exert no *control*

[26]To be clear, if the door is already broken down, this won't deter the burglar – viz., he will waltz right in and steal everything that he can.

[27]If you are sure that your door will break down in the event that you intervene to kick it, then this probability should be 1/1000.

over the burglar's selection.[28] This can readily be seen in the Figure 8 causal graph.

Figure 8: The Causal Structure of FD



$K = \{kick, \neg kick\}$

$D = \{down, \neg down\}$

$B = \{burgled, \neg burgled\}$

Here, if you update by conditionalization on your evidence that $D = down$ and maximize CEU while treating your options as $intervene_{kick}$ and $intervene_{\neg kick}$, then you will opt to kick down the door if you hate getting your stuff stolen enough. Thus like FMR2, FD demonstrates that modeling your options as interventions is not sufficient for delivering causal-decision-theoretic verdicts when choice is exotic. The problem is once more that of exotic colliders. By updating on a collider that is downstream from your choice, you can induce a correlation between your choice and something else that you care about. FMR2 is exhibit A; FD is exhibit B. Does this mean that the interventionist is up the creek without a paddle when choice is exotic?

[28]Though it's clear that causal-decision-theoretic reasoning speaks in favor of refraining from kicking when confronted with FD, it's less clear whether this is the right response. Maybe you should knock down the door so that the reason it breaks is that you break it, rather than that the burglar breaks it. Here, I take no stand on whether the causal decision theorist is right (in this case or any others), and leave further exploration of this issue to the science fiction writers of the world.

5 An Interventionist Solution

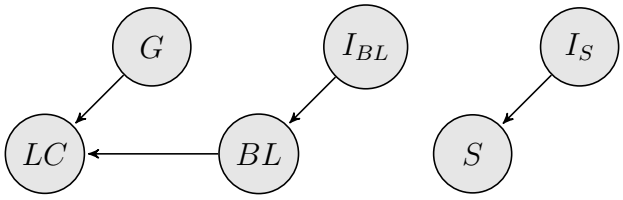
The problem for interventionists is that spurious (non-causal) correlations can arise between the agent's intervention and other states of affairs when the agent updates by conditionalization on evidence about the outcome of her choice, or, put differently, on variables that are causally downstream from the agent's options. In order to solve this problem, the interventionist must devise some new scheme for updating on exotic evidence.

Here, interventionists can take a page out of their old book. Before exotic choice reared its head, the trick to securing causal-decision-theoretic verdicts was to treat the conditioning event in CEU calculations as the intervention to act, rather than the act itself. We now know that this isn't sufficient for yielding causal-decision-theoretic verdicts in general, but this doesn't mean that the path to exotic causal-decision-theoretic verdicts isn't similar. In fact, as things turn out, the interventionist can once again secure causal-decision-theoretic verdicts by swapping out some conditioning event for an intervention — namely, by updating by conditionalization on the *intervention* to bring about the exotic evidence, rather than the exotic evidence itself. Consider FMR2 again.

If the interventionist maintains that the rationality of your choice depends on what maximizes CEU when you conditionalize on $intervene_{blacken}$ (rather than *blacken* itself), the dominance argument for smoking goes through successfully. This is because conditioning on I_{BL} effectively breaks the arrows between BL and its non-intervention parents (as depicted in Figure 9), and I_S and LC are therefore not correlated in the updated distribution. Intuitively, when you update on the intervention to blacken your lungs, you update on the fact that your lungs will

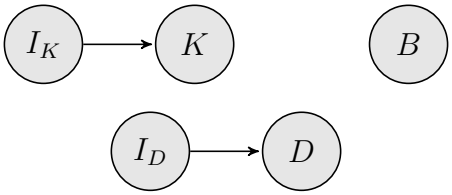
blacken for reasons that are independent from both your current choice and your genes. And of course, if you learn that your lungs will blacken for reasons that have nothing to do with your choice or genes, then opting to smoke is not even evidentially related to whether you will get lung cancer. This is because we cannot infer that it's more likely that your lungs will blacken because you have the gene than because you opt to smoke when we know that your lungs will blacken no matter how you choose and no matter what your genes are. Thus if you maximize CEU conditional on I_S and I_{BL}, you will judge that there is no reason not to smoke.

Figure 9: Interventionist Updating on Exotic Evidence in FMR2



The same goes for FD. As Figure 10 makes clear, when you conditionalize on $intervene_{down}$ rather than *down* itself, you update on the fact that your door is broken for reasons that are independent from both whether you kick and whether the burglar burgles your house. Thus there is no longer any reason — evidential or otherwise — to endure the mild pain of the kick.

Figure 10: Interventionist Updating on Exotic Evidence in FD



It is easy to see that this approach to updating on exotic evidence will always avoid the problem at hand. There is no risk of inducing a spurious correlation between the agent's intervention and any other variable by conditioning on some intervention(s) to bring about exotic evidence because interventions on variables that are causally downstream from the agent's intervention cannot themselves be causally downstream from the agent's intervention, and therefore cannot be colliders (or descendants of colliders) between the agent's intervention and anything else.[29] Thus while conditioning on a collider can induce a correlation between its causes, conditioning on the intervention to fix a collider cannot.

Still, despite this solution's ability to incorporate exotic evidence without inducing any spurious correlations, one might worry that it does not supply a sufficiently general updating procedure because there are some kinds of exotic evidence that cannot be modeled in terms of interventions. More specifically, one might worry that any context wherein the agent has *uncertain* exotic evidence about some effect of her choice poses a problem because interventions are typically defined such that intervening on a variable *determines* the value of the intervened upon variable.[30] Thus the concern would be that there is no straightforward way

[29] In the language of d-separation, interventions to make exotic evidence obtain cannot unblock any paths between the agent's intervention and any other variables in $\mathbf{V}$.

[30] An agent has uncertain exotic evidence when she acquires exotic evidence about some variable without learning the true value of that variable. There is already one case of exotic choice in the literature that seems to involve uncertain exotic evidence — namely, Egan's (2007) Alexandria case in which an agent has access to a time machine that can be used to go back to the

to use interventionist machinery when exotic evidence is uncertain.

Though there may be no *straightforward* way to use interventionist machinery in these contexts, there *are* interventionist methods that are useful here. Specifically, we can update on a variable that stands in the same causal relevance relations as a typical intervention variable — i.e., that occupies the same position in a DAG — but that is set up so that conditioning on its values do not *determine* the values of the intervened upon variable. Consider a version of FMR2 where the variable BL is fine-grained so that it has three values (instead of two) — namely, $\neg$*blackened*, *moderately blackened*, and *severely blackened*. If you learn that your lungs will blacken, but don't learn how much, then you should spread your confidence exclusively over *moderately blackened* and *severely blackened*, but it is intuitive that you shouldn't rule out either. Here, again, we cannot simply condition on the disjunction of *moderately blackened* and *severely blackened* because doing so will induce a spurious correlation between your choice and whether you get cancer, but we *can* condition on an exogenous *manipulation* of BL, M_{BL}, that is rigged up so that $P(\neg blackened|manipulate_{blackened}) = 0$.[31] This doesn't yet fully solve the problem

Library of Alexandria immediately before historical records show that it burned down. The agent contemplates whether she should send a fleet of fire trucks back to try to save the Library of Alexandria from burning down, or, alternatively, go back in time herself to steal a single treasured volume from the library before it burns down. Since the agent knows from the present that the library will actually burn down, but can't rule out that some volume will make its way to her in the future, Egan argues that the agent should opt to steal the volume. Hitchcock (2016) models this case in terms of a ternary outcome variable whose values are that zero volumes get saved, that one volume gets saved, and that every volume gets saved. Given this construal of the case, Egan's agent has exotic evidence that every volume will not get saved, but this qualifies as uncertain evidence because she is in no position to rule out either of the other two values of the variable.

[31]Here and in what follows, I refer to this kind of variable as a *manipulation* variable in order to flag its difference from a standard intervention variable. Unlike interventions, manipulations do not always break the causal influence of manipulated variables from their non-manipulation causal parents. For example, in the present context, M_{BL} should be construed such that conditioning on $manipulate_{blackened}$ does not totally sever the causal dependence of BL on S (and instead so that the extent to which your lungs blacken still depends on S).

since there are generally multiple ways to construe such a manipulation, and they can differ with respect to the resulting posterior (e.g., with respect to how your confidence should be spread across the two remaining values of BL).[32] But it does show that even when exotic evidence is uncertain, the interventionist can incorporate exotic evidence without inducing any spurious correlations by updating on an exogenous cause of the exotic evidence variable, rather than the exotic evidence variable itself. When the exogenous cause deterministically sets the value of the exotic evidence variable, there is but one posterior compatible with the prior and the independence constraints imposed by the DAG. But when the exogenous cause does *not* set the value of the exotic evidence variable, there are multiple coherent posteriors (or "manipulated distributions") that are compatible with the prior and the DAG, and further considerations must therefore be summoned to adjudicate between them (as well as between the various ways of construing the manipulation variable).[33]

[32]See Eberhardt and Scheines (2007) and Korb et al. (2004) for extensive discussion of non-standard kinds of manipulations.

[33]Though space limitations prevent me from arguing for a particular approach to selecting the posterior (or manipulated distribution) in contexts like these, Stephan Hartmann and I are working on developing an updating procedure that may prove useful here. Our basic idea is that we should opt for the posterior distribution that is most *conservative* in the sense that it minimizes some divergence (e.g., Kullback-Leibler) from the prior while satisfying some constraints that are imposed by the update. Here, the relevant constraints are, first, that it integrate the new unconditional probabilities for the exotic evidence variable (i.e., the probabilities supplied by the uncertain learning experience), and, second, that the unconditional probabilities for any parents of the exotic evidence variable remain at their prior values. (The CMC can be used to justify this latter constraint since its truth implies that the parents of the exotic evidence variable must be probabilistically independent from any exogenous manipulation to the exotic evidence variable.)

6 The Principle of the Matter

We now have a simple two-step recipe for using interventionist machinery to secure causal-decision-theoretic verdicts no matter whether the choice at hand is exotic. First, as usual, model the agent's options as interventions and require that agents maximize CEU. Second, when the agent has exotic evidence, insist that the agent maximize CEU relative to the probability distribution that results from *intervening* or *manipulating* to bring about the exotic evidence (rather than the probability distribution that results from conditioning on the truth of the exotic evidence itself).

So that's a recipe, but is it principled? The standard interventionist can argue that it's reasonable to treat the agent's options as interventions because agents must represent their choices as causally autonomous from the factors under consideration (even when they actually aren't).[34] The idea here is, first, that an agent makes a genuine choice only when their choice is uncaused by the factors under consideration, and, second, that the standards of rationality apply only when the agent makes a genuine choice, no matter how unlikely it is that the agent is making a genuine choice.[35] But what about the second step? There *prima facie* seems to be no good argument for the interventionist approach to updating on exotic evidence. After all, there are many Bayesian arguments for conditioning on the content of your evidence, and the interventionist approach to exotic choice seems to flout this norm insofar as the agent is supposed to maximize CEU relative to the

[34]Here, and in what follows, I drop explicit mention of non-intervention manipulations for the sake of convenience.

[35]See Meek and Glymour (1994) and Stern (2018) for more extensive discussion of interventionist treatments of choice.

distribution that is updated on the intervention to bring about the exotic evidence (again, rather than the content of the exotic evidence itself).[36] Consider FMR2. The content of your evidence is that your lungs will blacken, not that your lungs will blacken for reasons that are independent from your current choice. But the interventionist approach to exotic choice speaks in favor of updating on the latter, not the former. What gives?

There may be multiple ways for the interventionist to justify this approach, but the most straightforward (at least that comes to mind) involves maintaining that agents not only must represent their choices as causally autonomous, but also must represent their choices as evidentially irrelevant to (or autonomous from) any variables about which they have evidence at the time of choice. When choice is not exotic, this evidential autonomy is secured by modeling the agent as intervening since the CMC entails that the intervention to act is uncorrelated with any variable about which the agent could have non-exotic evidence.[37] But when choice is exotic, this evidential autonomy is *not* secured by modeling the agent as intervening (since the intervention to act is typically correlated with its effects), but *is* secured by additionally modeling the agent's exotic evidence in terms of interventions or manipulations since, e.g., $intervene_{blacken}$ is not correlated with $intervene_{smoke}$. The reason, then, that you should update on $intervene_{blacken}$ as you deliberate

[36] It is worth noting that this may constitute the beginnings of a novel argument against causal decision theory. Before the problem of exotic choice reared its head, it seemed as though causal-decision-theoretic verdicts could be readily attained in a Bayesian framework simply by modeling the agent's options as interventions and updating on evidence in a standard Bayesian fashion. But now that the problem of exotic colliders is on the table, it turns out that if we want to attain causal-decision-theoretic verdicts, we may have to abandon one of the two pillars of Bayesianism — i.e., Bayesian conditionalization.

[37] This likewise explains why the interventionist updating procedure developed here applies *only* when evidence is exotic. When evidence is not exotic, conditioning on the evidence itself is consistent with making a genuine choice because treating the agent's choice as causally autonomous implies that the agent's choice is evidentially autonomous from such evidence.

is that your choice counts as genuine only if your choice is independent from any variables about which you have evidence, and it is independent from these variables only when your evidence that your lungs will blacken is construed as the intervention that your lungs will blacken.[38] To be clear, this does not mean that it's epistemically rational to be certain that your lungs will blacken for reasons that are independent from the choice that you're making. After all, for all you know, your lungs may blacken because you will settle on smoking. It's just that as you deliberate, you need not worry about the worlds in which your lungs blacken for these reasons because these are worlds in which your choice is *not* evidentially autonomous from your evidence.[39]

This conception of choice as evidentially autonomous has plenty of precedent in the philosophical literature. For example, both Joyce (2007) and Velleman (1989) argue that "a deliberating agent who regards herself as free need not proportion her beliefs about her own acts to the antecedent evidence that she has for thinking that she will perform them." Similarly, Ramsey (1929) writes the following:[40]

> "The past, we think, is settled; if this means more than that it is past, it might mean that it is settled for us, that nothing now could change our opinion for us of any past event. But that is plainly untrue. What is true is this, that any possible present volition of ours is (for us)

[38] A variable qualifies as independent (in the relevant sense) from the agent's choice when it is probabilistically independent from the agent's choice before conditioning on any value of it.

[39] It may be natural to think that when choice is exotic, there is retro-causation from the exotic evidence variable to your choice (since your choice is informed by the value of the exotic evidence variable). By the interventionist's lights, worlds in which causal hypotheses like these are realized are irrelevant to what's rational because the agent is not causally autonomous from the factors under consideration and therefore is not making a genuine choice according to the interventionist.

[40] See Liu and Price (forthcoming) for extensive discussion of the relationship between Ramsey's (1929) view and Joyce's (2007) view.

> irrelevant to any past event. To another (or to ourselves in the future) it can serve as a sign of the past, but to us now what we do affects only the probability of the future."

Ramsey's conception of choice is at least very closely related to the conception of choice that justifies the interventionist approach to updating on exotic evidence.[41] According to Ramsey, as we deliberate, we must bracket any possibilities according to which our "volition" is probabilistically relevant to anything that we take to be settled, even though these possibilities may be relevant to determining what our volition makes likely as we don't deliberate.[42] When we treat only the past as settled (as we do when choice is not exotic), the probabilistic irrelevance between our present volitions and the past is secured by modeling our volitions as interventions. But when choice is exotic, we treat some aspect of the future as settled (in the relevant sense), and Ramsey's line of thought therefore seems to suggest that we should treat our present volitions as irrelevant to these future aspects. Again, this does not square with treating our options as interventions and updating on the exotic evidence itself, but it does square with characterizing *both* our options and our exotic evidence in terms of interventions.

[41] This conception of choice is also closely related to Levi's (1989) view that "deliberation crowds out prediction" and Spohn's (1977) view that there should be "no probabilities for acts." See Stern (2018) for explicit discussion of these views in the context of graphical causal models and Liu and Price (forthcoming) for explicit discussion of the relationship between these conceptions of choice and Ramsey's.

[42] Ramsey addresses the conflict between what is rational for predictive purposes and what matters for rational choice by distinguishing the probabilistic judgments from our deliberating selves from those who aren't deliberating (including our future selves). This isn't the exact line of reasoning developed above since there it was maintained that the standards of practical rationality apply only when the agent is making a genuine choice, even when there is rational pressure for everyone (including the deliberating agent) to regard their choice as evidentially constrained, and therefore not genuine. But Ramsey still develops the core insight that we should treat the agent's choice or "volition" as evidentially autonomous while determining the practically rational choice.

Still, there is a noticeable asymmetry between the interventionist's method for securing evidential autonomy in the case of ordinary choice and the case of exotic choice that calls for explanation. By modeling the agent as intervening, the interventionist ensures that the agent's choice is uncorrelated with any variable that the agent does not regard as causally downstream from her action, no matter whether the agent has evidence about any such variable, and thereby guarantees that the agent's choice is independent from anything about which the agent *could* have non-exotic evidence. For example, in BTD, when we model your choice whether to smoke as an intervention, your choice is rendered independent from both G and LC regardless of whether you know the status of either variable. But the interventionist updating scheme for exotic evidence is not like this insofar that it targets only those variables about which you *do* have evidence. For example, in FMR2, if we never acquired the exotic evidence that our lungs would blacken, then the dependence between our choice and the state of our lungs would have remained intact. Thus it is only when we receive the news from the oracle that our lungs will blacken that we intervene to break this dependence. What explains why we should construe genuine choice so that it requires independence from any factors about which the agent *could* have non-exotic evidence while only requiring independence from those causally downstream (exotic) factors about which the agent *does* have evidence?

This asymmetry can be justified by arguing that genuine choice requires *both* causal and evidential autonomy. In the interventionist framework, when choice is not exotic, the former implies the latter (because the CMC implies that causally autonomous interventions are not correlated with anything that the agent could have non-exotic evidence about). But when choice is exotic, the interventionist

must take additional measures to secure evidential autonomy — i.e., she must deploy the updating procedure developed here in order to secure independence from the factors about which the agent actually has evidence. Since there are plenty of defenders of both kinds of autonomy in the literature, this may constitute a sufficiently principled defense of the asymmetry (especially since the combination of the two yields decision-theoretic verdicts that causal decision theorists find intuitive).[43] But it is worth noting that there is another way to respond to the asymmetry. Namely, one could argue, first, that exotic decision-making contexts reveal the central importance of evidential autonomy as a constraint on genuine choice (insofar as we must ensure that evidential autonomy is secured when treating choice as causally autonomous doesn't suffice), and, second, that this provides reason to replace (rather than supplement) the standard interventionist's causal autonomy constraint with the evidential autonomy constraint.[44] This response is a non-starter here since it does not vindicate causal-decision-theoretic reasoning.[45] But it may be attractive to those who are not committed causal decision theorists.[46]

Either way, our foray into the realm of exotic choice has revealed something potentially deep about the conception of agency that underwrites causal-decision-theoretic reasoning (at least when viewed through the lens of graphical causal models). In order to secure causal-decision-theoretic verdicts in every possible decision-making context, the interventionist must maintain not only that genuine

[43]See Stern (2018) for discussion of both autonomy constraints in the context of interventionist causal models.

[44]See Stern (2018) for discussion of this possibility.

[45]Consider its application to BTD. In the event that you lack evidence about G or LC, your choice could qualify as evidentially autonomous even when it's correlated with either variable. Thus if you maximize CEU, the causal-decision-theoretic dominance argument for smoking will not go through successfully.

[46]This may be of a piece with Huw Price's general line of response to Newcomb problems. See Price (2012).

choices are causally autonomous from the variables under consideration, but also that they are evidentially autonomous. Were our focus directed exclusively toward non-exotic choice, we would miss this since the former implies the latter when choice is not exotic. Thus it is only by focusing on exotic choice that we draw out the central role that considerations of evidential autonomy play in causal-decision-theoretic reasoning.

7 Ordinary Choice

Before concluding, it is worth taking stock of why it matters what choices we should make when choice is exotic. The primary reason to care about exotic choice is that it provides a nice testing ground for theories of agency by offering a context in which to probe their decision-theoretic implications. We've specifically learned that if causal-decision-theoretic responses to exotic Newcomb problems like FMR2 are right, then the ordinary interventionist conception of choice should be replaced with the Ramseyan conception that fares better more generally. But I suspect that this won't quash every doubt about the importance of this guide. Some readers who are less preoccupied with issues in the philosophy of action may wonder whether there are any actual decision-making contexts in which the advice developed here is pertinent. In this section, I argue that there are three kinds of actual decision-making context in which this guide is relevant.

First, following Rabinowicz (2009), it is important to note that exotic choice is of genuine practical relevance when agents mistakenly take their choices to be exotic (regardless of whether they actually are). We know all too well that there are actual agents who fit this description since we have all encountered people

who believe that they have seen or know the future. For example, no matter whether oracles actually exist, we all know real people who put their faith in oracles (or psychics, or crystal balls, or whatever), and these people can easily come to believe that they have evidence about some variable(s) that they take to be causally downstream from their choice(s). You and I may think that such people are mistaken, but the mistaken need advice, too.[47] Thus it would be undesirable for these people's choices to lay beyond the domain of our best theories of rational choice. Or put differently, if it's possible to develop a decision theory that provides an evaluative standard for choices made given these mistakes, then we should. This interventionist's guide to exotic choice fills this lacuna.

Second, as Lewis (1982) points out, exotic decision-making contexts are clearly logically possible since there are logical possibilities that enable the acquisition of exotic evidence (e.g., time travel and precognition). Even if these possibilities have not been actualized at this point in time, they could be someday.[48] Thus our

[47]Decision theories are standardly thought to provide advice to agents who believe all sorts of silly things. For example, even flat-earthers can look to decision theory for advice about how to choose given their unjustified credences about the state of our planet. Why should agents with unjustified credences about the future be treated any differently?

[48]One might object that this line of reasoning is flawed because it's *conceptually* impossible to make an exotic choice — i.e., that no agent can have evidence about the outcome of her choice as she makes it (because the nature of choice precludes this possibility). There are two things worth saying here in reply. First, even if this is right, it doesn't erase the need for a decision theory that handles exotic decision-making contexts since there could be agents who misapply the concept of 'choice' so that they mistakenly take themselves to make exotic choices. Second, there is a sense in which this line of objection is not at odds with the advice developed here. I have argued that the interventionist can deliver causal-decision-theoretic verdicts when choice is exotic by insisting that the agent condition on the intervention to bring about the agent's exotic evidence, rather than the exotic evidence itself. Moreover, I have argued that this may be principled because it preserves the evidential autonomy of choice. Since the intervention to bring about the exotic evidence is *not* exotic itself (because the intervention is not causally downstream from the agent's choice), there is a sense in which the agent must treat her evidence as non-exotic (and therefore a sense in which her choice should be regarded as non-exotic) in order for the evidential autonomy of her choice to be preserved — i.e., she must update on the non-exotic intervention, rather than the exotic evidence itself. Thanks to an anonymous referee for pressing this objection.

decision theories had better provide an evaluative standard for these decisions in order to count as sufficiently complete.

Third, and perhaps most interestingly for those who worry about the practical relevance of exotic choice, there may be contexts in which choice isn't genuinely exotic, but in which the advice developed here is nevertheless pertinent. These are contexts in which actual decision-makers use an impoverished language as they make decisions, relative to which genuinely ordinary evidence is naturally modeled as exotic. This happens when an agent's evidence about the past and/or present is such that it cannot be directly accounted for in the agent's impoverished language (because there is no event in the algebra on which to condition), but nevertheless justifies an inference about the value of some variable that is causally downstream from the agent's choice.

Suppose, for example, that you learn from a recent journal article that some causal hypothesis consistent with Figure 7 is correct, and that you also have recently learned from a medical test that your lungs are already black. Is there any reason not to smoke?

As we evaluate your choice, it is natural to start with a probability distribution that captures what you learned from the statistical model in the journal article, and then to update this probability distribution to reflect the known particulars of your case. Since the statistical model may be defined over just those variables depicted in Figure 7, this can mean starting with a prior that is defined over just these variables.[49] Relative to such a prior, there is no way to directly represent

[49]One might argue that the ideally rational agent would not deploy such an impoverished language, and would instead work with a language that is expressive enough to allow for conditioning precisely on what is learned (i.e., that the agent's lungs are *currently* black).This is fine as far as it goes, but actual decision-makers sometimes make choices relative to probability distributions that are defined over impoverished algebras. Moreover, this phenomenon is ubiquitous

your ordinary evidence that your lungs are black in terms of conditioning on an event. But you may also know that blackened lungs stay black, and thus may need to update your prior in correspondence with your knowledge that your lungs will be black.[50] If you update by conditionalization on this fact, then, as with FMR2, you will judge that there are cancer-related reasons not to smoke. But if you instead update on the intervention to make your lungs black, you capture the sense in which there is intuitively no cancer-related reason not to smoke, given that your lungs have no hope, and that the causal influence of S over LC is entirely causally mediated by the state of your lungs.[51]

Just as you need to update your best knowledge of a causal system with the particulars of your case as you decide whether to smoke, decision-makers confronted with weighty choices must do the same. For example, policy-makers must update their best causal understanding of economic systems with the particulars of their specific economies as they decide how to regulate currency, doctors must update their best causal understanding of the human body with the particulars of their patients' bodies as they decide whether to perform surgery, and so on and so forth. Frequently, our best understanding of such causal systems is based on

in contexts like the one described here — i.e., when agents' choices are informed by statistical models that are themselves defined over impoverished languages. In these contexts, agents often rely on the model not just as they determine their confidence in the states of affairs over which their confidence is defined, but also as they define (or "carve up") the states of affairs themselves. Here, our focus is on how agents should choose, *given* the impoverished language over which their priors are defined.

[50] If you do not believe that the current state of your lungs determines their future state, then it would be inappropriate to update BL in this case.

[51] Interestingly, in circumstances like these, updating on the intervention is the epistemically rational update. If you know the value of some variable that is downstream from your choice, and don't know it because you know the value of non-intervention variables in $\mathbf{V}$ that determine the causally downstream variable, then provided that $\mathbf{V}$ is causally sufficient (in the sense that it satisfies the interventionist's requirement that any common causes of any variables in $\mathbf{V}$ are themselves included in $\mathbf{V}$), then you know the value for reasons that are accurately modeled in terms of an intervention.

statistical models that are given over impoverished languages that don't include events for the precise evidence that we have about some particular case. Thus a lot rides on how we should update our best understandings of causal systems when we have case-specific knowledge, including case-specific knowledge about variables that are causally downstream from the intervention under consideration.

8 Conclusion

This concludes our interventionist guide to exotic choice. The interventionist methods developed here improve upon their predecessors not only insofar as they help us to evaluate what's rational when choice is exotic, but also insofar as they help us (i) to better understand the conception of agency that vindicates causal-decision-theoretic reasoning when viewed through the lens of graphical causal models, and (ii) to make rational choices in ordinary contexts wherein we have evidence about some variable that we represent as causally downstream from the choice at hand. There may still be some issues to work out — e.g., whether it's possible to propose novel counterexamples to causal decision theory in the context of exotic choice[52]—

[52]It is plausible that modeling exotic evidence in terms of interventions sometimes yields counterintuitive verdicts when updating on the intervention additionally breaks a dependence between the exotic evidence variable and a variable that is *not* causally downstream from the agent's choice. In these cases, updating on the intervention to bring about the exotic evidence (rather than the exotic evidence itself) can break the dependence between the exotic evidence variable and some background circumstances that are relevant to what the agent's choice is likely to cause. This bears a striking similarity to what goes on in Egan's (2007) putative (non-exotic) counterexamples to causal decision theory, where modeling the agent's choice as an intervention breaks the dependence between the action variable and some background circumstances that are relevant to what the agent's action will cause. (Hitchcock (2016) discusses this aspect of Egan's "psychopath" case in great detail.) For the purposes of this paper, I grant Egan that causal-decision-theoretic reasoning yields the verdicts that he finds counter-intuitive in these cases, and only aim to develop a guide to exotic choice that captures causal-decision-theoretic reasoning. But this is definitely an issue that deserves attention in the future. Just as Joyce (2012) develops his version of causal decision theory so that it does not deliver counter-intuitive verdicts to Egan's

but the primary aim of this paper has been to use interventionist causal models to capture causal-decision-theoretic reasoning when choice is exotic, no matter the ultimate fate of causal decision theory.[53]

non-exotic counterexamples, interventionists may be able to develop their guide to updating on exotic evidence (and the underlying conception of agency that justifies it) so that dependencies between exotic evidence and variables that are not causally downstream from the agent's choice are not broken. Thanks to Sebastian Krug and an anonymous referee for pressing this point.

[53]For helpful discussion and comments, I am very grateful to Thomas Blanchard, Catrin Campbell-Moore, Benjamin Eva, Malcolm Forster, Melissa Fusco, Clark Glymour, Stephan Hartmann, Daniel Hausman, Christopher Hitchcock, Sebastian Krug, Ben Levinstein, Huw Price, Shanna Slank, Julia Staffel, Rush Stewart, two anonymous referees, an editor of this journal, and the audiences at the 2018 Causes, Norms, and Decisions Workshop in Hanover, Germany and the 2019 Formal Epistemology Workshop in Turin, Italy.

References

Bales, A. (2016). The Pauper's Problem: Chance, Foreknowledge, and Causal Decision Theory. *Philosophical Studies* 173: 1497–1516.

Bradley, R. (2018). *Decision Theory with a Human Face*. Cambridge: Cambridge University Press.

Easwaran, K. (forthcoming). A Classification of Newcomb Problems and Decision Theories. *Synthese*.

Eberhardt, F. and Scheines, R. (2007). Interventions and Causal Inference. *Philosophy of Science* 74: 981–995.

Egan, A. (2007). Some Counterexamples to Causal Decision Theory. *Philosophical Review* 116: 93–114.

Gibbard, A. and Harper, W. (1978). Counterfactuals and Two Kinds of Expected Utility. In C. Hooker, J. Leach and E. McClennen (ed.) *Foundations and Applications of Decision Theory*. Dordrecht: Riedel, pp. 125–162.

Hitchcock, C. (2016). Conditioning, Intervening, and Decision. *Synthese* 193: 1157–1176.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Joyce, J. (2007). Are Newcomb Problems Really Decisions? *Synthese* 156: 537–562.

Joyce, J. (2012). Ratifiability and Stability in Causal Decision Theory. *Synthese* 187: 123–145.

Korb, K., Hope, L., Nicholson, A., and Axnick, K. (2004). Varieties of Causal Intervention. In *Proceedings of the Pacific Rim International Conference on AI*. New York: Springer.

Levi, I. (1989). Rationality, Prediction, and Autonomous Choice. In *The Covenant of Reason: Rationality and the Commitments of Thought*. Cambridge: Cambridge University Press, pp. 19–39.

Lewis, D. (1982). Letter to Wlodek Rabinowicz.

Lewis, D. (1986). Causal Decision Theory. In *Philosophical Papers: Volume 2*. Oxford: Oxford University Press, pp. 214–240.

Liu, Y. and Price, H. (forthcoming). Ramsey and Joyce on Deliberation and Prediction. *Synthese*.

Meek, G. and Glymour, C. (1994). Conditioning and Intervening. *The British Journal for the Philosophy of Science* 45: 1001–1021.

Pearl, J. (1993). Comment: Graphical Models, Causality, and Intervention. *Statistical Science* 8: 266–269.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. 2nd edition. Cambridge: Cambridge University Press.

Pearl, J. (2016). Lord's Paradox Revisited – (Oh Lord! Kumbaya!). *Journal of Causal Inference* 4.

Price, H. (2012). Causation, Chance, and the Rational Significance of Supernatural Evidence. *Philosophical Review* 121: 483–538.

Rabinowicz, W. (2009). Letter from Long Ago: On Causal Decision Theory and Centered Chances. In J. Lars-Goräin, J. Österberg, and R. Śliwiński (eds.) *Logic, Ethics, and all that Jazz: Essays in Honour of Jordan Howard Sobel.* Uppsala, pp. 247–273.

Ramsey, F. P. (1929). General Propositions and Causality. In D. H. Mellor (ed.) *Foundations: Essays in Philosophy, Logic, Mathematics and Economics.* London: Routledge, pp. 133–151.

Reichenbach, H. (1956). *The Direction of Time.* Berkeley, CA: University of California Press.

Sobel, J. (1980). *Probability, Chance and Choice.* Unpublished.

Spohn, W. (1977). Where Luce and Krantz Do Really Generalize Savage's Decision Model. *Erkenntnis* 11:113–134.

Stern, R. (2017). Interventionist Decision Theory. *Synthese* 194: 4133–4153.

Stern, R. (2018). Diagnosing Newcomb's Problem with Causal Graphs. In A. Ahmed (ed.) *Newcomb's Problem (Classical Philosophical Arguments).* Cambridge: Cambridge University Press, pp. 201–220.

Velleman, J. D. (1989). Epistemic Freedom. *Pacific Philosophical Quarterly* 70: 73–97.