1 Why did they do that?

Ross Shachter* and David Heckerman†

1.1 Abstract

Judea Pearl argues that people as well as machines with artificial intelligence must use causal reasoning to make decisions and to explain or justify those decisions. We wholeheartedly agree with Judea on this point, but show that the ability to identify available alternatives and the ability to express preferences are also necessary for making and explaining decisions. We briefly review the basic principles of decision theory, showing how these three abilities come together in decision making. We illustrate these principles with examples including Judea's incisive depiction of the story of Adam and Eve.

1.2 Introduction

At the start of "The Book of Why," Judea Pearl places us in the Garden of Eden [Pearl and Mackenzie 2018]. When God asked Adam "Have you been eating of the tree I forbade you to eat?" Adam replied "It was the woman you put with me; she gave me the fruit and I ate it." Eve added "The serpent tempted me, and I ate" [Jones 1971]. Judea notes that instead of responding "Yes" or "No" to God's questions, they gave excuses, explanations of *why* they ate the fruit, attempting to shift the blame from themselves to others. They knew that they disobeyed God and they hoped they could avoid the consequences that God declared, "of the tree of the knowledge of good and evil you are not to eat, for on the day you eat of it you shall most surely die."

Judea argues that asking "Why did they do that?" is a natural question and argues that causal reasoning, the second level of his Ladder of Causation, is a necessary tool to answer this question. We wholeheartedly agree with these points [Heckerman and Shachter 1995]. In particular, in order to know why Adam and Eve ate the fruit, we need to know how they thought about the causal consequences of their possible actions. Judea also argues that machinery for causal reasoning is a necessary component of true Artificial Intelligence. We agree with this point as well. That said, we use this article to argue that there are

* Department of Management Science and Engineering, Stanford University, Stanford, CA
† Amazon, Seattle, WA

two additional machineries needed to answer the question of "Why did I do that?" and further argue that these machineries are equally important components of a true Artificial Intelligence. The first additional component is the ability to identify available alternatives—that is, what an entity can do in a given situation. The second component is the ability of the entity to express preferences over the possible outcomes.

These three components for answering the question "Why did I do that?" are precisely the key aspects of Decision Theory [Blackwell and Girshick 1954, Howard 1970, Raiffa 1968, Ramsey 1926, Savage 1954, von Neumann and Morgenstern 1947]. It is from this perspective that we explore the story of the Garden of Eden to illustrate the approach. (We do not intend to convey any theological insight.) Before we return to their decision and a description of how they could use Decision Theory to give a more complete explanation of why they ate the fruit, let's consider a simple decision problem and how it is represented by Decision Theory.

1.3 Some examples

Consider whether to pay for parking if we expect to be staying for only a short time to run an errand. This situation is represented by the decision tree shown in Fig 1.1.

We can choose to pay for the parking or we can choose to not pay for it. These are our alternatives. If our violation is detected then we will receive a ticket and have to pay a fine, but if it goes undetected then we will have parked for free. For many people, parking for free is the best outcome and paying a fine is the worst outcome, so there is some tradeoff involving the cost of parking, the cost of the fine, and the probability that our violation will be detected.

While we don't want to get much into the details of Decision Theory, we note that, once the ingredients of alternatives, probabilities over possible outcomes (from causal reasoning), and preferences are specified, the theory offers a prescription to act. In this simple example, we should pay for parking if

$$P\{detection\}(\text{Cost of Fine}) > (\text{Cost of Parking}).$$

Another important feature of Decision Theory is that different decision makers can have different alternatives, probabilities, and preferences. For example, a decision maker may know there is free parking nearby and add that as an alternative, may be uncertain about the cost of the fine if they fear their car might be towed, or prefer not to park without paying.

The identification of alternatives, the assessment of probabilities through causal reasoning, and the assessment of preferences are all important ingredients to making a decision, and answering the question "Why did I do that?" In a recent CACM article, Judea talks

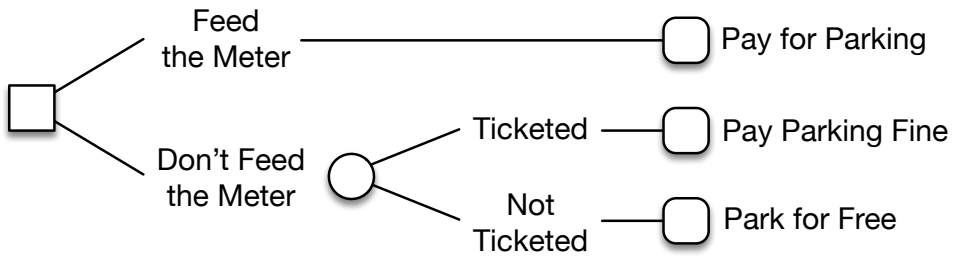

Figure 1.1 We can choose to feed the meter and pay for parking, or choose not to feed it and then either pay a fine or park for free, depending on whether our car is ticketed.

about how causal reasoning can help someone who asks "I am about to quit my job, but should I?" [Pearl 2019]. We agree that causal reasoning is an important component to this task, but knowing alternatives and preferences are equally important. We can't imagine how causal analysis could answer this person's question without knowing his preferences and the alternatives available to him. Is he considering another job or going back to school? Is the issue his work situation, his family, his health, or his commute?

1.4 Back to the Garden of Eden

Now let us return to Adam and Eve's decision, viewing it from their perspective *before* they ate the fruit. (In this story, we are in the very unusual situation where eating the fruit actually changes who they are, potentially changing their ability to identify alternatives and to reason causally, and potentially changing their preferences. So, we must be careful to look at their decision before they ate the fruit.) Their situation is represented by the decision tree shown in Figure 1.2. Their alternatives are to eat the fruit or to not eat it. Thinking about the consequences about their possible actions—that is, thinking causally—they reason that, if they don't eat the fruit, everything will remain as it has been. If they eat the fruit and God finds out, God has said that they shall most surely die, presumably by God's hand. In contrast, the serpent has said "No! You will not die! God knows in fact that on the day you eat it your eyes will be opened and you will be like gods, knowing good and evil." So they reason that there are two possibilities: either God will find out and they will die, or he won't find out and they will get to enjoy the fruit and any knowledge that comes with it. "The woman saw that the tree was good to eat and pleasing to the eye, and that it was desirable for the knowledge that it could give." Therefore, enjoying the fruit is the best outcome and dying is the worst outcome, and they need to think about how likely it is that God will find out and how desirable the fruit must be to take that risk.

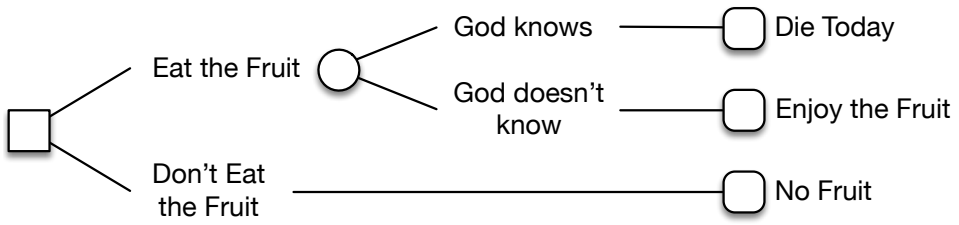

Figure 1.2 Adam and Eve can choose to eat the fruit and then either die or live having enjoyed the fruit, depending on whether God finds out, or they can choose not to eat the fruit and never get to enjoy it.

We know that they ate the fruit. Perhaps they had considered other alternatives—for example, pleading with God if God were to find out. Or perhaps they considered other outcomes—for example, that they might not die if God finds out. However, the Bible is silent on these other possibilities, so we leave the representation of their decision—"Why did they do that?"—as shown in the decision tree in Figure 1.2.

1.5 Decision Theory and Decision Analysis

There are other interesting aspects surrounding the story of Adam and Eve, but before we return to these, let us consider Decision Theory in more detail. Decision Theory is a *normative* theory for decision making. That is, its principles follow from a small set of axioms that a decision maker should follow. For example, one of the axioms is that a decision maker's preferences over the possible outcomes are totally ordered. If not, for example if A is preferred to B, B is preferred to C, and C is preferred to A, then a third party could extract money from the decision maker, by getting him to pay for a preferred outcome over another, until he is back with the outcome he started with, but with less money. Decision theory is not a descriptive theory of decision making—that is, it does not accurately describe *how* people actually make decisions.

Prospect theory is such a theory, attempting to explain and characterize the processes people use in decision making, incorporating their biases and heuristics.[Kahneman and Tversky 2006] Prospect theory would be an appropriate theory for an Artificial Intelligence if our goal were merely to simulate human decision makers. However, it seems more compelling to consider the normative Decision Theory developed to explain and characterize the processes humans might want to use for important decisions.

Decision Analysis, developed by Ron Howard and Howard Raiffa, is a discipline centered on the application of Decision Theory to real decisions in practice. One of the fundamental distinctions in decision analysis is between the quality of a decision made and

the quality of the resulting outcome. Although it is commonplace in our society to judge a decision by the outcome, they are quite different, as good decisions can have bad outcomes and bad decisions can have good outcomes. We use Decision Analysis to help us make a decision *before* we act, improving the quality of our decision. Using Decision Analysis only to explain our decision *later* leaves most of its benefits on the table.

Unfortunately, it is when a decision maker experiences a bad outcome that they are most likely to ask themselves (again) "Why did I do that?" That leads to the question "What would have happened had I made another choice instead?" This is often how we learn to make better decisions, and it is the type of counterfactual reasoning on the third level of Judea's Ladder of Causation.

Another important notion from Decision Analysis is that, in any given situation, a pure Do(X=x) alternative might not be available, and the decision maker must choose from the limited set of realistic alternatives. For example, consider the decision made by a patient and her physician about which treatment she should receive for a serious disease. Ideally, she would like the alternative Do(disease=false) with no side effects, but this is rarely available as an alternative.

An extreme example of side effects can be found in the classic horror story of "The Monkey's Paw" [Jacobs 1902]. In that story, each owner of a monkey's paw was granted three wishes. Although their wishes were fulfilled, it was not done so as the wisher had intended, such as when his wish for money was fulfilled by the death of his son leading to an insurance payout.

Indeed, identifying realistic alternatives can be difficult. Nonetheless, this identification is a necessary component for answering the question "Why did I do that?" and a necessary component for true Artificial Intelligence. In humans, the process of identifying alternatives seems to arise naturally from the experience of free will. Interestingly, recent experiments in neurobiology [Soon et al. 2008] and physics [Proietti et al. 2019], cast doubts on whether humans actually have free will or merely perceive that we do. Nonetheless, many believe that at least the perception of free will is necessary for a true Artificial Intelligence. This remains an open question.

1.6 Back again in the Garden of Eden

Now consider when God asks Adam and Eve "Have you been eating of the tree I forbade you to eat?" Adam blames Eve, who looks back at her decision making and uses causal reasoning to recognize that the serpent's words led her to both lower her probability that God would find out and increase her preference for the fruit, together causing her to choose to eat the fruit. Eve summarizes this to God, saying "The serpent tempted me, and I ate."

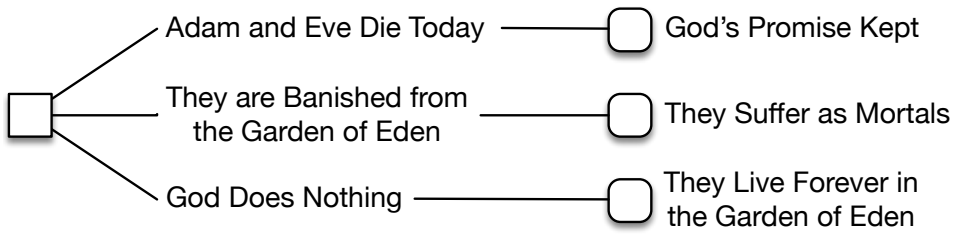

Figure 1.3 Knowing they have eaten the fruit, God has three choices: either Adam and Eve die as God promised, they are banished from the Garden of Eden to a life of suffering as mortals, or they will live forever in the Garden of Eden.

Of course, Eve leaves out the important details of the temptation, which she must have surmised—with her new knowledge of good and evil—that God would not appreciate. Interestingly, the serpent's statement turned out to be true, if misleading: Adam and Eve did not die that day, as God had threatened, and they indeed gained the knowledge of good and evil.

1.7 Conclusion: God's decision

We conclude by thinking about the decision that God made after confronting Adam and Eve from the perspective of Decision Theory. When God created them, we learn "Now both of them were naked, the man and his wife, but they felt no shame in front of each other." However, when God found them in the Garden of Eden after they ate the fruit, Adam says "I was afraid because I was naked, so I hid." God's causal model kicks in: "Who told you that you were naked? . . . Have you been eating of the tree I forbade you to eat?"

It might not have taken any inference, as some believe that God is all knowing and can predict the future as well. Those people believe that God knew all along that Adam Eve would eat the fruit. However, others believe that when God created Adam and Eve he endowed them with curiosity and free will. God could only control their behaviors imperfectly and indirectly through punishment and reward. And to do so perfectly, God would have needed to understand their alternatives, their causal reasoning, and their preferences!

Once God confirmed with Adam and Eve that they ate the fruit, there was a decision to be made about their punishment. If they died that day, then God's promise would have been fulfilled. If nothing happened, then God had fears: "See, the man has become like one of us, with his knowledge of good and evil. He must not be allowed to stretch his hand out next and pick from the tree of life also, and eat some and live for ever." It would seem that God had decided that they should die.

However, God introduced a new alternative of banishing them to a life of suffering as mortal humans, as shown in Fig 1.3. But why? God makes it clear that it was neither forgiveness nor mercy, for he cursed them and all of their descendants with lives of pain and suffering. Did God merely threaten death to dissuade them from eating the fruit knowing that otherwise they could not resist the temptation? If only we knew God's available alternatives, causal reasoning, and preferences well enough to understand "Why did God do that?"

Bibliography

D. Blackwell and M. A. Girshick. 1954. *Theory of Games and Statistical Decisions*. Wiley, New York.

D. Heckerman and R. Shachter. 1995. Decision-theoretic foundations for causal reasoning. *J. Artificial Intel. Res.*, 3: 405–430.

R. A. Howard. 1970. Decision analysis: Perspectives on inference, decision, and experimentation. *Proc. IEEE*, 58(5): 632–643.

W. W. Jacobs. 1902. The Monkey's Paw. *Harper's Monthly*, 105: 634–639.

A. Jones. 1971. *The Jerusalem Bible*. Doubleday, Garden City, NY.

D. Kahneman and A. Tversky. 2006. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2): 263–292. ISSN 00129682. DOI: 10.2307/1914185.

J. Pearl. 2019. The Seven Tools of Causal Inference, with Reflections on Machine Learning. *Commun. ACM*, 62(3): 54–60.

J. Pearl and D. Mackenzie. 2018. *The Book of Why*. Basic Books, New York.

M. Proietti, A. Pickston, F. Graffitti, P. Barrow, D. Kundys, C. Branciard, M. Ringbauer, and A. Fedrizzi, 2019. Experimental rejection of observer-independence in the quantum world. http://arxiv.org/abs/1902.05080.

H. Raiffa. 1968. *Decision analysis: introductory lectures on choices under uncertainty*. Addison-Wesley.

F. P. Ramsey. 1926. Truth and probability. *Philosophical papers*, pp. 52–94.

L. J. Savage. 1954. *The Foundations of Statistics*. Wiley, New York.

C. S. Soon, M. Brass, H. J. Heinze, and J. D. Haynes. 2008. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, 11(5): 543–545. ISSN 10976256. DOI: 10.1038/nn.2112.

J. von Neumann and O. Morgenstern. 1947. *Theory of Games and Economic Behavior*, 2nd. Princeton University Press, Princeton, NJ.