Prisoners of Prophecy: Freedom and Foreknowledge in the *Dune* Series

> Prophecy and prescience – How can they be put to the test in the face of the unanswered questions? Consider: How much is actual prediction of the "waveform" (as Muad'Dib referred to his vision-image) and how much is the prophet shaping the future to fit the prophecy? … Does the prophet see the future or does he see a line of weakness, a fault or cleavage that he may shatter with words or decisions as a diamond-cutter shatters his gem with a blow of a knife?
>
> "Private Reflections on Muad'Dib" by the Princess Irulan

Among the special abilities of characters in the Duniverse, prescience – the ability to see into the past, present, and future – is both the most isolating and the most strange. Guild Navigators use spice to acquire the foresight needed for safe interstellar travel. Paul attains an extraordinary level of prescience, even compared to the Navigators. His sister Alia as well as his children, Leto and Ghanima, possess these abilities to a phenomenal degree. Leto's foresight and its consequences dominate the *Dune* books from *God Emperor* to *Chapterhouse*. Some other characters, like Mother Superior Darwi Odrade and Siona Atreides, have relatively limited prescience.

Why is prescience isolating for its possessor? As Paul finds when talking to Stilgar, he cannot explain it to those without the power: it's beyond their comprehension.[1] The deceptive strangeness of prescience in *Dune* is typical of Herbert's ideas.[2] At first, it seems like magic, but Herbert is taking something familiar to our modern perspective – the ability to make predictions – and making it strange. We can't prophesize like Paul, but we can "see"

the future, sometimes with great accuracy, using scientific knowledge, common-sense reasoning, consulting experts, and so on. The ancient Babylonians were able to systematically predict astronomical events, but contemporary astrophysicists can forecast distant events beyond the Babylonians' wildest dreams.

Herbert describes the prescience of characters like Paul as a hyper-awareness of possibilities and probabilities given certain choices, rather than being able to examine a fixed future. The course of time is not entirely predetermined. Paul and Leto II must confront a nightmare in which all their possible choices involve horrific events. The only paths they can see for avoiding human stagnation and extinction involve tyrannically supressing humans for millennia and then unleashing a cultural reaction against that despotism. Their prescience is not magical, but instead an unimaginable ability to anticipate the outcomes of present choices. That's an amplification of what we can already do – for instance, we can model the global climate and estimate outcomes of various environmental policy choices, but after that, we must choose how to act given this information. Of course, Herbert doesn't give us all the details of how prescience works; if that knowledge were public, characters in the Duniverse wouldn't see the powerful predictors as a messiah or a God Emperor.

There are two areas of real-world research especially relevant to the questions of prescience that Herbert explores. First, there is decision theory, which combines ideas in philosophy, economics, psychology, and other fields. Decision theorists study how we make choices and which choices are rational. Second, there is the philosophy of action, which is about the puzzles and insights surrounding our abilities to choose, to will, and to succumb to temptation, among other actions. Philosophy of action intertwines with psychology,

neurology, the philosophy of causation, and other areas. For our purposes, it's the area in which questions about the relation between freedom and foreknowledge fall.

The Prisoner's Dilemma

Common sense suggests that prescience should help us live together better. After all, you're faced with uncertainty if you come across a stranger on a dark night in an unfamiliar country. Preparing for trouble (like going into a fighting stance) might make them nervous, escalate the situation, and lead into a spiral of violence. If you knew that the stranger wouldn't harm you, then you wouldn't have to risk this reaction and its consequences. Many wars would never have been declared if their outcomes had been foreseen.

However, decision theorists can show that the foresight of Paul and Leto II creates traps that ordinary people might escape. One of these is the "Prisoner's Dilemma." To understand this trap, let's first make it clear what decision theorists mean by "rationality": your beliefs and desires, whatever they might be, should fit together coherently. A simple example: you shouldn't believe both of two statements that contradict each other. So if you believe that you have ten fingers, it's irrational to also believe that you have eleven fingers.

This sense of "rationality" differs from our everyday sense of the word. It's "formal" rationality, which decision theory understands through the logical and mathematical relations among a person's beliefs and preferences, rather than what those beliefs and desires are about. In contrast, we often talk about "rational beliefs," "irrational desires," and so on. For example, if someone desires the short-term pleasures of a drug knowing about the long-term

suffering that it will cause, then we might say that their desire is "irrational." But this judgment is not something we reach by looking at formal rationality.

For the Prisoner's Dilemma, this formal sense of "rationality" entails this: if you regard all the possible results of an action A_1 as more desirable than the possible results of another action A_2, then you should choose A_1 rather than A_2. So, if Leto II believes that the Golden Path will *inevitably* be better than any alternative, then it is rational for him to lead humanity in that direction.

The Prisoner's Dilemma can be interpreted in different ways, but the basic logic is always the same. It is what decision theorists call a "game," again a word with a technical meaning. A game is any interaction between decision-makers involving strategies. A "strategy" is a plan of action. Poker is a game, but battles, trade negotiations, and even romantic dates can also be seen as "games." In the Prisoner's Dilemma, we assume that the interacting "players" are formally rational. We also assume the players each have only two possible actions; that they know the outcomes of these actions; and that they know that both players are rational.

The Prisoner's Dilemma is an imaginary interaction between "prisoners," but it actually might make more sense with prescient characters from the Duniverse. The trap is that rational strategic behavior by the players leads to what decision theorists call a "suboptimal" outcome: consequences that aren't the best possible results. Imagine that Paul is interacting with the prescient being developed by the Bene Tleilax in one meeting that will be their only interaction. This prescient being is mentioned in *Dune Messiah* and called "Thallo" in *Paul of Dune*. Departing from the books, let's assume that Thallo has identical abilities to

Paul. These two prescient beings can choose between fighting and cooperating. They'll make their choices simultaneously. Furthermore, with their prescience, they can predict the consequences of their choices. If they both fight, *Galactic Conflict* occurs with unpredictable ultimate results. If either chooses to fight while the other tries to cooperate, then the player who chooses to fight will win. Finally, if they both choose to cooperate, then they rule jointly, and we assume that, despite the uncertainties involved, they prefer *Joint Rule* to *Galactic Conflict*. We can lay out the possible choices and consequences using what decision theorists call a "payoff matrix":

	Thallo Fights	Thallo Cooperates
Paul Fights	*Galactic Conflict*	*Paul Rules*
Paul	*Thallo Rules*	*Joint Rule*

To create a trap, all we need is to assume that Paul and Thallo have these preferences:

(a) Paul prefers *Paul Rules* to *Joint Rule*, both of those to *Galactic Conflict*, and *Galactic Conflict* to *Thallo Rules*.

(b) Thallo prefers *Thallo Rules* to *Joint Rule*, both of those to *Galactic Conflict*, and *Galactic Conflict* to *Paul Rules*.

Notice how both players prefer *Joint Rule* to *Galactic Conflict*; both realize that *Galactic Conflict* is not the best-case scenario. Now, decision theory says that a rational player will

make a decision that is one of their best choices, given other players' decisions. When all players reason this way, we have a "Nash equilibrium," named after the economist John Forbes Nash (1928–2015) who developed the idea. In the Prisoner's Dilemma, there is only one Nash equilibrium. We can find it by elimination. It can't be *Joint Rule*, because if Paul is cooperating, then Thallo better satisfies his preferences by fighting, and vice versa. It can't be *Paul Rules*, because if Paul is fighting, then Thallo prefers fighting as well. And it can't be *Thallo Rules*, because if Thallo is fighting, then Paul prefers fighting as well. Therefore, the Nash equilibrium is *Galactic Conflict* – what Paul and Thallo would choose if they are deciding rationally.

The Prisoner's Dilemma paradoxically shows how rational choices can lead to suboptimal outcomes. As strange as this might seem, it happens in the real world. For example, imagine a very crowded bar. Suppose that the optimum outcome is that everyone talks at a moderate level: the different groups in the bar can't hear each other perfectly, but still fairly well. However, it could be individually rational for each group to raise their voices slightly to hear each other better, then slightly more when the other groups do the same thing, until it's almost impossible for anybody to hear each other. Again, rationality leads to a suboptimum outcome. Economists call such situations "market failures," but they also happen in situations that we don't normally call "markets," like politics.[3]

Foreknowledge and the Iterated Prisoner's Dilemma

It's possible to modify the Prisoner's Dilemma to avoid a market failure. Imagine now that Paul and Thallo must repeatedly choose between fighting and cooperating – the "iterated" Prisoner's Dilemma. If they don't know how many times they'll play this game, then

choosing to cooperate can be rational, because this makes it possible to avoid the other player choosing to punish you in the future.

One rational strategy is called *tit-for-tat*: cooperate in the first game, and in future games only choose to fight ("defect") if the other player fought you in the previous game. Additionally, *tit-for-tat* requires cooperating if the other player cooperated in the previous game. *Tit-for-tat* swiftly punishes defecting, but also rewards the other player for switching to cooperation. If both Paul and Thallo start out with this strategy and stick to it, then they will start with *Joint Rule*, and keep choosing it. Also, if Paul chooses *Tit-for-Tat*, then there are a lot of other strategies that Thallo could choose where they will ultimately end up cooperating. Iterated interactions can offer one escape from the trap of rational suboptimality.

Yet iterated Prisoner's Dilemmas can still trap the prescient. Imagine that there are exactly ten Prisoner's Dilemma interactions between Paul and Thallo. Suppose that, before the tenth game, both of them can foresee with certainty that this game is their last together. Assuming that their preferences fit the pattern (a) and (b) above, it's rational for both of them to choose to fight in the tenth game, because they can't punish each other. Now, imagine that the end of the interactions is foreseeable just before the ninth game. Both Paul and Thallo know that fighting will be the rational choice for both of them in the tenth game; there's no incentive to cooperate together in the ninth game, because they can't be rewarded for cooperation in the tenth game. The same reasoning works whether the end becomes foreseeable on the ninth game, the eighth game, and so on, up to and including the first game. Therefore, if Paul and Thallo have powers of prescience that enable them to identify the end of their interactions, they will be trapped in rational suboptimality. So prescience is not always a gift; it can be a trap. Since our own knowledge of the future is just a weaker version

of what Paul and Thallo can do, it's possible that our own foresight can lead us into similar disasters.

Here is one of the many places where philosophers can add something insightful to decision theory. Common sense suggests that it is rational to achieve optimal outcomes. Does the Prisoner's Dilemma imply a contradiction between decision theory and common sense?

I've been using "rationality" in the formal sense, implying the coherence of preferences and beliefs but not what our beliefs are about. Philosophers have developed stronger senses of "rationality," though. Aristotle (384–322 BCE) argued that being rational requires having the right kinds of preferences: a rational person desires to perform their family roles well, to be a good friend, to be a good citizen, and so on. A rational person, for Aristotle, prefers to be a good person, and so rational people will prefer not to benefit at the expense of others, as occurs in *Paul Rules* and *Thallo Rules*. If both Paul and Thallo know that they have these preferences, then they know that they will both prefer *Joint Rule* to ruling, and thus avoid a Prisoner's Dilemma. Similarly, the philosopher Immanuel Kant (1724–1804) argued that rationality requires acting according to a rule that everyone else could follow in the same situation (similar to the Golden Rule: do unto others as you would have them do unto you). This would exclude the outcomes *Paul Rules* and *Thallo Rules*, so Kantian rationality would achieve the optimal outcome of mutual cooperation.

If Aristotle and Kant are closer to our everyday concept of rationality, then there may be no genuine conflict between decision theory and common sense: both can be useful for different purposes. We just need precision in how we define "rationality."

Prediction and Rationality

Adding to the problems with prescience, it can paradoxically cause conflict between two intuitive rules for rational decision-making:

> *Maximization*: A rational decision-maker acts to maximize the sum of how much they prefer (or dislike) each foreseeable outcome of each action, weighted by (a) the strength of their preference and (b) how probable they think each outcome will be, given that action. This sum is called "expected utility."

> *Dominance*: If one action A1 is better than another action A2 under some foreseeable circumstance and A2 is not better than A1 under any foreseeable circumstance, then it is rational to do A1 rather than A2.

In a puzzle called Newcomb's Paradox, these principles seem to conflict. Suppose that you are playing a game with a Guild Navigator. Their prescience is imperfect, but you know they can reliably predict your choices. The Navigator's aides have put money in two boxes. You can see into Box A, which has $1000 inside, while you can't see into Box B. You are offered two possible choices, One-Box (choosing just Box B) or Two-Box (choosing Box A and Box B). If the Navigator predicted that you will pick Two-Box, then they left Box B empty. If the Navigator predicted that you will pick One-Box, then they will have put $1,000,000 in it. You won't know their prediction until after you've chosen.

	Navigator Predicts One-	Navigator Predicts Two-
One-Box	$1,000,000	$0
Two-Box	$1,001,000	$1000

You can choose one of two strategies, One-Box and Two-Box, and let's assume that the strength of your preference corresponds to the amount of money you win. The principle of *Maximization* suggests that you should choose One-Box, because you know that the Navigator is an extremely reliable predictor, and the expected payoff of One-Box is about $1,000,000. However, Two-Box always gives a higher payoff than One-Box: $1,001,000 vs. $1,000,000 in the first column, $1000 vs. $0 in the second column. Therefore, the principle of *Dominance* recommends Two-Box.

Decision theorists disagree about which choice is most rational. One approach is Causal Decision Theory (CDT), in which the aim of rational action is not to maximize an action's expected utility, but instead the utility of the action's consequences. The Navigator isn't directly observing the future, so their choices aren't *caused* by your decision. So reasoning in terms of *Maximization* doesn't work: you can't cause the Navigator to predict that you will One-Box, and so you can't cause your winnings to increase by choosing it. In contrast to CDT, Evidential Decision Theory (EDT) says to maximize your expected utility *given your evidence*, and this is achieved by choosing One-Box, since it gives you evidence that the Navigator predicted this choice.

Neither supporters of CDT nor of EDT have yet found any definitive arguments for

their preferred theory. Since both CDT and EDT seem plausible, we might worry that these paradoxes reveal a fundamental incoherence in our concept of formal rationality. However, in recent years, a third position has developed, which argues that when we specify the details of such paradoxes precisely (more precisely than I've done) then EDT and CDT give the same recommendations.[4] If this view is successful, then there might be little or no real disagreement between EDT and CDT. In general, little details in a scenario's formulation can make a big difference in decision theory.

Self-knowledge and Freedom

One of the most curious things about prescient beings like Paul or Leto II is that they know things about *their own* future behavior. Yet, in this sense, we all have some foreknowledge. I can predict that I'm not going to become an early riser tomorrow, although I might desire that. A psychology student learns more about her own future behavior as she studies. In neither case is the foreknowledge absolutely certain: it's conceivable, though extremely unlikely, that I might knock my head on a door tonight and become an early riser. Paul and Leto II really only differ from us in their self-knowledge's extent and reliability.

It might seem that, as knowledge of our future behavior expands, our sense of our own freedom contracts. Leto II can see that there are certain plans that he cannot complete: they are psychologically impossible for him, because they are too ruthless or painful. The knowledge itself doesn't restrict him, but should he feel less free? Assuming we want to feel free, is this kind of self-knowledge a curse?

Stuart Hampshire (1914–2004) argued that self-knowledge tends to have the opposite

effect – that it makes us freer.[5] If Leto II knows that he can't follow a plan, then he can avoid making commitments he can't fulfil. In this way, he'll be able to accurately plan his life, and this is just part of what it is to be free. Hampshire also argues that if we believe that we have *any* choice in a situation, an increase in our self-knowledge from science or other sources doesn't stand in the way of our sense of freedom. No matter how much we know in such a situation, we can always step back and ask, "If these are the actions that are possible based on my psychological features, which action will I choose?"

Hampshire's view of action builds on Francis Bacon's (1561–1626) argument for scientific research: knowledge is power. According to Hampshire, self-knowledge is also power. But what if we lived in a theocracy where we'd be imprisoned if we discovered certain facts? Even in that case, it's the theocrats who are responsible for our loss of freedom, not the self-knowledge. Self-knowledge from prescience could be uncomfortable for Paul and Leto II (they might discover nasty truths about themselves, like a potential for great cruelty), but it's actually the opposite of a trap.

What about learning that we never had any freedom? Hampshire surprisingly says that we would see that we do not decide at all – in fact, we would learn that we do not think, because thinking presupposes making decisions. Can't we predict what others will decide? Yes, Hampshire says, but until we make a decision we can only learn the boundaries of our choices, not what we will choose. And to "know what we will do" is to decide, not to predict.

At least in the Duniverse, prescience can work with this view. Prescience is a vastly expanded understanding of possible and probable consequences given information and choices, not the observation of a fixed future. Having examined their possible futures, Paul

and Leto II can still ask, "What shall I do?" This sort of prescience seems empowering, not entrapping.

We started with the question, "Is prescience a trap?" It definitely can be, as the Prisoner's Dilemma reveals. However, Duniverse prescience need not rob Paul or Leto II of their sense of freedom, if Hampshire is right. Does self-knowledge suggest that our concept of rationality is incoherent? That depends on the answer to puzzles like Newcomb's Paradox and other exciting debates between philosophers, economists, psychologists, and others. Fortunately, none of us are prescient enough to foresee their results.

---

[1] *DM*, Ace, 60.

[2] Consider also the Voice and the imprinting powers of the Bene Gesserit and the Honored Matres. Herbert takes familiar abilities (verbal manipulation, sexual manipulation) and considers their improvement by over 20,000 years of refinement.

[3] David Friedman, *The Machinery of Freedom* (Scotts Valley, CA: CreateSpace Independent Publishing Platform, 2014), ch. 53.

[4] Wolfgang Spohn, "Reversing 30 Years of Discussion: Why Causal Decision Theorists Should One-box," *Synthese* 187 (2012), 95–122; David H. Wolpert, and Gregor Benford, "The Lesson of Newcomb's Paradox," *Synthese* 190 (2013), 1637–1646.

[5] Stuart Hampshire, *Thought and Action* (Notre Dame, IN: Notre Dame University Press, 1959) and *Freedom of the Individual* (Princeton, NJ: Princeton University Press, 1975).