

[Working draft. Please do not circulate or cite without author's permission]

Gyges' Choice

Rationality and visibility

Josiah Ober

Chapter 1 of *The Greeks and the Rational* (book in progress, provisional title)

Draft of 2019.0902 Word count (w/ notes and bibliography): 20,400

ABSTRACT: In *Republic* book 2 Plato's Glaucon offers a reformulation, via the thought experiment of "Gyges and the ring," of an ancient Greek folk theory of practical reasoning – a coherent account of human deliberation, choice, and action that deductively yields explanations and predictions about behavior. The theory bears a family resemblance to contemporary (20th/21st c.) normative and descriptive theories of rational choice, in its motivating intuitions and in the habit of illustrating abstractions through fanciful stories and historical narratives. It assumed that a rational agent would have ordered preferences over outcomes and beliefs about the state of the world, and that such an agent would act accordingly, based on expectations of desire satisfaction. Plato's thought experiment is based on an anecdote in Herodotus' *Histories* (book 1) describing how Gyges became king of Lydia. Plato's account concerns unconstrained choice and is modeled by a simple decision tree. Herodotus' earlier story, lacking the conceit of an invisibility ring, emphasized the constraints on individual desire satisfaction that arise from others' choices; it can be modeled as a three-player game in extensive form.

Pages 1-6 are from an Introduction chapter, in progress.

To speak of practical reason as a "discovery" invites comparison to Molière's disingenuous M. Jourdain in *Le bourgeois gentilhomme* (1670: act 2, scene 6): "Well, what do you know about that! These forty years now I've been speaking in prose without knowing it!"¹ Molière's renowned send-up of the pretensions of 17th century French intellectuals reminds us that the Greeks (and others) were quite capable of employing reason in pursuit of their goals long before the instrumental uses of rationality became a subject of theoretical inquiry. But, as I hope to show, the systematic theorization of practical reason in fifth and fourth-century Greece counts as a discovery, and that discovery has a history. It was only one aspect of a wider intellectual revolution, but it had specifiable effects when the theory was integrated into Greek social and political thought, and when it was put into practice. Those effects are especially evident in democratic Athens.

In the following pages I address two questions: What was discovered? And what were the implications of the discovery for Greek thought and practice? The short answer is that instrumental rationality was recognized as a powerful and versatile tool for making choices under conditions of constraint and uncertainty. The use of that tool was refined as a craft (*politikê technê*), notably by the so-called Sophists. It was characterized as a social, political, and ethical problem by historians and by Socratic philosophers. And, finally, it was integrated into the philosophical systems of Plato, Aristotle, and their successors as an essential foundation for moral reasoning. I make no claim here for the *uniqueness, priority, or superiority* of the Greek discovery: Roughly contemporary and highly sophisticated traditions of reasoning from means to ends arose, for example, in China and India.² Yet, by the same token, Greek thought about practical reason can be studied as a distinctive, well-documented, and influential intellectual tradition.

Rationality ancient and modern

This book explores ancient Greek thinking about how individuals and groups make decisions and act on them. The book's thesis is that Greek writers, beginning with Homer and Hesiod and continuing through Plato and Aristotle (and beyond), concerned themselves with intellectually challenging issues related to the practice of reasoning from desire through means to ends. These issues – including self-interest as a source of individual motivation, the possibility of coherent and effective collective action, and the strategic calculations involved in taking into account chance and the likely behavior of others when choosing among options – lie at the heart of contemporary (20th- and 21st-century) formal theories of decision and games. When applied to social questions, economic behavior, or politics, government, and international relations, these approaches are summed up under the rubrics of rational choice, strategic behavior, and positive political theory.

Formal decision and game theories are predicated on explaining choice and action in terms of instrumental rationality, understood as a consistent ordering of preferences and the maximization of expected utility. I argue that ancient Greek thinkers and lawmakers anticipated some of the assumptions about desire, belief, and expectation (although not the mathematical expression) that underpin contemporary formal theories. Greek writers and institutional designers also recognized that there were limits to human rationality; they knew that real choice-makers could not be reduced to mechanistic calculators of optimal outcomes. Yet they typically saw these limits as adjustments to or deviations from a fundamentally rational approach to decision-making under conditions of uncertainty, rather than as a non-rational alternative.³

This line of thought, if it is on the right track, has significant consequences for how we think about the development of ancient Greek civilization, the history of

thought about practical reasoning, and perhaps about human rationality itself. The tendency of the last two generations of scholarship on classical Greece has been to emphasize how *different* ancient Greek culture, thought, and social practices were from the forms of rationality characteristic of modernity.⁴ Meanwhile, contemporary game theorists and social choice theorists (e.g. Binmore 2007; Sen 2017) tend to trace the origins of their fields back to philosophical and mathematical thought of the 18th century – e.g. to Hume, Rousseau, and Condorcet.

The question then arises: does the “rationality of choice” capture a deep (if necessarily partial) truth about cognition and human behavior generally? Or is it only (if at all) a contingent feature of some people, inhabiting certain roles in certain contemporary societies? Of persons who think and act as they do because they have been subjected to a culture of science and to economic conditions that are unique to modernity? Insofar as we, today, are instrumentally rational, is it because we are human? Or because we are modern?⁵ Even if one accepts, as I do, that modernity has put its stamp on expressions of humanity, these questions may seem important. This book is addressed to those who think that they are.

It is, I suppose, uncontroversial to say that reasoning about how individuals and groups make choices provided part of the intellectual infrastructure for Greek writers, including Plato, Aristotle, Herodotus, Thucydides, Xenophon, Polybius, Plutarch, and the dramatists. They aimed, in the first instance, at contributing to (or revolutionizing) existing bodies of work in the domains of history, moral and political philosophy, comedy, tragedy, and so on. There was no ancient Greek discipline or literary genre that corresponds to the modern fields of decision and game theory, or to the modern disciplines of economics and political science. But the central claims of this book are, first, that ancient Greek writers knew, developed, and criticized ideas about rationality, choice, and action in ways that were sophisticated, systematic, and well worth closer investigation. And, next, that those ideas were widely operationalized in social norms and formal rules. Various ancient Greek political institutions are best understood as the result of rational choices and bargains. As such, they can readily be modeled by simple strategic games.

In order to explore Greek thought and practice on the rationality of choice, I employ the standard methods of classical philology and history, offering a series of close readings of Greek texts, literary and documentary. I analyze the vocabulary of, for example, choice, constraint, advantage, profit, and likelihood used in those texts.⁶ I draw attention to metaphors (e.g. path or road) and allegorical tales (e.g. about Gyges the Lydian) that Greek writers employed to illustrate the relationship between the micro-foundations of individual decisions and social behavior. I also seek to show, by reference to decision trees and simple games, illustrated in normal or extensive form, that the authors of the relevant Greek texts and institutional arrangements aimed at posing and answering some of the questions that are raised

by contemporary choice literature. Although I introduce a number of Greek terms and some of the terminology employed by positive political theorists, I do not assume that the reader has prior knowledge of either classical Greek or game theory. Technical matters and passages in the original Greek are relegated to the endnotes.

One glaring difference between antiquity and modernity will be immediately apparent: Greek writers and legislators concerned with the rationality of choice lacked the mathematical formalization and algebraic expression that exemplifies contemporary formal theories of decision and games. Even when we restrict ourselves to disciplinary domains centered on ethics, politics, and social organization (leaving aside the domain of mathematics and its associated technical fields), the mathematical formalization of choice theory has obvious benefits: It is a uniquely powerful and rigorous kind of abstraction and systematization. As such, formal theory makes it possible to treat hard problems that lack intuitive solutions, to compare seemingly dissimilar cases by using standard methods, and to incorporate and analyze quantitative data.

I do not pretend that the Greeks anticipated those essential features and virtues of contemporary formal theory. But, insofar as the basic premises of formal theories of decision and choice can be explained in ordinary language and simple illustrations, without algebra (as introductory texts by prominent game theorists demonstrate), it remains possible that some of the core intuitions about motivation, rationality, and social behavior on which formal theory is predicated would be accessible to a culture lacking the distinctive methods and results of mathematical reasoning that made contemporary formal theory possible and that are its defining features. And, taking the next step, that such a culture might conjoin informal understandings of how the various intuitions cohere into a theoretical paradigm useful for explaining (and perhaps even predicting) phenomena that otherwise remained opaque. And, finally, such a culture might develop ways of thinking about practical reason that are sophisticated, powerful and not readily captured by the mathematical tools employed by contemporary choice theorists. This book contends that the ancient Greeks did just that and that their achievement in this area merits the attention of anyone concerned with rationality.⁷

Ancient Greek folk theories of practical reasoning

In the *Memorabilia* (3.9.4) the Athenian polymath Xenophon quotes Socrates as follows: “For I think that all persons choose (*proairoumenous*), out of what is available to them, what they think is most advantageous to themselves (*sumphorôtata autois*), and they do this.” In this succinct sentence, offered as an rebuttal to the proposition that persons who know what they ought to do and yet do something else (i.e. the problem of *akrasia*) are “clever and willful,” Xenophon’s

Socrates offers a concise description of a presumptively universally-available process of practical reasoning that results in action (“they do this”).⁸

Socrates’ account of choice-making in the passage cited is universalizing in its claim that a standard approach to reasoning about advantage is available to “all persons.” That approach is deliberative, in that it is a process of advance reasoning leading to a choice (*proairoumenous*). The process of choosing is based on agents’ relevant beliefs (*oiontai*) and is carried out under contingent conditions of constraint (“out of what is available to them”). The goal of the agents is maximization of advantage (*sumphorôtata*) to the agents themselves (*autois*).

Socrates does not imply (or at least need not be read as implying) that a process of clear-headed reasoning is the unique approach to choice-making employed by all persons in all circumstances of willed action.⁹ It may be that some (or indeed many) human actions are based on momentary impulse. Beliefs about what is most advantageous may arise from emotion or illusion. The point, then, is not that we can be sure that Xenophon’s Socrates has given us the germ of an account aimed at explaining a purely rational mental process underpinning all human behavior. But he has given us an account of the process of reasoning from available means to desired ends that we are to suppose is universally available to people, and pervasive enough to bear on the important question of whether or not people act out of weakness of will (*akrasia*).

The comment assimilates what, in Plato’s dialogues, we learn is Socrates’ own highly distinctive approach to choosing among possible courses of action to the process employed by ordinary people. Plato’s reader will have learned that Socrates’ understanding of what it takes for a person to make a choice that can properly be regarded as advantageous is demanding. It requires an assessment of “what is available,” a mode of thinking, a conception of advantage, and a knowledge of oneself that are beyond the ken of most people. In the passage cited above, Xenophon’s Socrates abstracts from the special kinds of knowledge and reasoning that Plato’s Socrates supposes are required in order to choose *well*.

In the Xenophon passage, the method of choosing among options employed both by Socrates and other people, and the relationship between their choices, beliefs, and actions, are imagined as being relevantly similar enough to claim that there is an approach to reasoning about advantage that is common to “all persons.” Although Socrates is often thought to have broken decisively with the background intellectual culture of his age, and most decisively with the thought of the Sophists, the *Memorabilia* passage cited above is included in a recent, authoritative edition of the fragments of the Greek Sophists by André Laks and Glenn Most. Laks and Most posit that, based on his interests and argumentative methods, Socrates can be seen as “an idiosyncratic Athenian ‘sophist’ (2016: 4-5, quote: 293). I propose that Xenophon has put in Socrates’ mouth a concise statement of a “Folk Theory” of

practical reasoning. That is, Xenophon's Socrates was not offering an original observation. While giving it a paradoxical spin, he was in agreement with an account of how people tend to choose to act that was (1) widely known among (at least) classical-era Athenian intellectuals, (2) could not be claimed as original by any one author, (3) was sufficiently coherent and analytic to count as a kind of theory.¹⁰

The Folk Theory accounts for behavior by reference to motivations and beliefs. It holds that a rational person will go for whatever she believes is best for herself, among the options she supposes are open to her.¹¹ It is a *theory* insofar as it is an axiomatic, internally coherent account of human deliberation, choice, and action. In Plato's philosophically sophisticated restatement, considered below (especially chapters 1, 2) the theory is normative (specifying how, to count as fully rational, an agent ought to act) and it deductively yields explanations and predictions that can be tested against hypotheticals and observable phenomena. The original "background" folk theory to which, as I suppose, Xenophon's Socrates referred, was descriptive in that it aimed at explaining some part of observed behavior. It was also normative in that it specified how a rational agent ought to behave. As I will seek to show in the course of this book, it was pervasive in classical and post-classical traditions of social thought and assumed by designers of certain formal institutions. While associated with the fifth-century Sophists, the folk theory antedates them and it cannot be traced back to any single source.

We cannot inquire of an ordinary Greek, living in the age of Socrates or Xenophon, whether he or she had any sort of theory of choice and action. But we do have many ancient Greek texts in multiple genres – including histories, biographies, forensic speeches, tragedies and comedies, political essays, and technical manuals – that shed light on background cultural assumptions concerning reasoning about choices. As I will hope to demonstrate in this and the following chapters, recognizing the cultural pervasiveness of a background "folk theory of practical reasoning," and tracing some of the variations on the background theory developed by Greek writers concerned with social and political thought and practice, can help us to make better sense of some well-known (and some less well-known) Greek texts and institutions. Moreover, juxtaposing variants of the ancient folk theory to modern theories of choice may help humanists and citizens to make better use of some techniques drawn from contemporary decision theory and game theory. If the argument of this book goes through, those techniques can be redeployed as analytic tools for exploring the ancient history of certain ideas often thought to be uniquely modern. The same tools might also be adapted for practicing an approach to citizenship often thought to be uniquely ancient.

Place-holder:

FT(B) Evidence for the background folk theory: tragedy, comedy, statements in the orators, gnomai, etc. The role of gods in the background theory.

FT(C) Folk theory reformulated in fifth century BCE in several variants. Will consider the well-known Thrasymachean variant below. But btw the background and Thrasymachean variants was what we can call conventionalist variant.

Examples include: Anonymous Iamblich, Lycophron in Aristotle Pol. Book 3

Note that Aristotle rejects Socratic take on akrasia – we return to this in ch. 8.

Chapter 1. Gyges' Choice.

1.1 Ancient folk theories and modern theories of rational choice

The ancient Greek folk theory of practical reasoning, in its original “background” form and in the more explicitly theoretical variants developed by fifth- and fourth-century BCE Greek writers, bears, as I will argue in the course of the book, a family resemblance to the contemporary theory of rational choice. The ancient theory, like the contemporary theory can, I believe, also be understood as a normative and (at least in aspiration) descriptive account of motivation, belief, and action. The “family” to which I suppose the ancient folk theory and modern rational choice theory belong may be thought of as the moral psychology associated, in the history of philosophy, with David Hume. Theories in this family understand choice as a psychological process involving both desire and belief. The agent’s motivating desires or preferences (conative states) are taken as being separate from, and not, themselves, derived from or motivated by the agent’s beliefs about the world (cognitive states). Once a preference has been formed, belief goes to work in calculating the availability and probability of desired outcomes. The agent chooses accordingly, going for the most-desired option among those believed to be available.¹²

Insofar as the family resemblance of the ancient folk theory to modern theories of rational choice actually does point to a shared approach to motivation and action, ancient Greek texts that applied the folk theory to real-world and hypothetical behavior and institutions may be illuminated by adapting some of the tools of modern decision theory and game theory, each of which takes the rationality of choice-makers either as a premise or a testable hypothesis. The ancient folk theory lacks the mathematical infrastructure and algebraic expression characteristic of these modern choice theories. Yet, so I will argue, some of the intuitions underpinning the ancient and modern theories are relevantly similar, as are the habits of illustrating the theory with fanciful stories and historical narratives, and testing it against the observed behavior of individuals, groups, and states.

In common with modern choice theory, the ancient Greek folk theory, as it was reformulated by Plato in the *Republic*, abstracted from ordinary human choice-

makers. It imagined in their stead hypothetical agents freed from certain of the cognitive limitations and social constraints that affect the choices of ordinary persons. Thus, like modern choice theory, the reconfigured ancient folk theory was *both* normative (concerned with how a fully rational individual ought to act) *and* descriptive (concerned with explaining why people act as they are observed to do). And like modern choice theory, the reconfigured folk theory was predicated on the assumption that a rational individual will act according to his desires and his beliefs. That is to say that he (1) has ordered preferences over outcomes;¹³ (2) has coherent beliefs about the relevant state of the world; and (3) acts accordingly, based on his expectations of preference satisfaction. Expectations are in turn predicated on the perceived likelihood of certain things coming to pass.

In the ancient and modern theories alike, after deliberation – getting clear about his preferences and beliefs relevant to a given situation, by thinking matters through and/or by discussing the situation with others – the rational individual forms a judgment about possible outcomes. He chooses the course of action that, having done the necessary calculations, he expects will lead to the best available outcome – that is, the one he thinks is most advantageous (profitable, beneficial, satisfactory) to himself.¹⁴ Under certain circumstances, the same process of rational choice may be attributed to social groups and to states, acting as collective agents. Thus, we may say that if subjective value (understood as an agent’s satisfaction with outcomes) is conditioned by his *expectations* regarding preference satisfaction (the *likelihood* of getting the best *available* outcome), rationality can be defined as behavior that maximizes expected subjective value – that is, per Xenophon’s Socrates (above), going for the *most* advantageous available outcome.

Such behavior may be roughly described as *self-interested*. Per Xenophon’s Socrates: what people go for is what they regard as most advantageous *to themselves*. Even when the agent’s highest-ranked preference is the good of others, getting that outcome is to her own advantage in the sense of desire satisfaction. Her rationality is *instrumental* insofar as the end is chosen because it is thought to be advantageous to the agent in the sense of satisfying a desire of the agent. It is not, therefore, chosen specifically for its own sake, as something that is choice-worthy as a good in itself, irrespective of its expected value for the choosing agent.¹⁵

The ancient Greek account of rational choice-making differs from modern choice theoretical accounts of “expected utility maximization” in that ancient writers did *not* suppose that agents “do the math” by multiplying the quantified probability of gaining an outcome (expectation) against the deduced desirability of that outcome (preference strength).¹⁶ Each of those alternative outcomes may have different probabilities of fulfillment and different levels of desirability. Doing the math would enable the agent to arrive at a sum of expected utility (potentially measured in some pseudo-currency, like “utils”) that could be directly compared to

the sum of expected utility attached to other available outcomes. As we will see (chapter 6), because the the ancient theory lacks this mathematical feature, it fails to offer the rational agent guidance in marginal cases, for example when there is a tiny chance of a highly desired or dreaded outcome.

The basic ancient and modern processes of choosing among available alternatives, are, however, relevantly similar in their underlying premises. Each is based on maximization of expected value to the agent. The difference is that the ancient method of calculation is qualitative. Although Greek theorists were very interest in the problem of applying scientific measurement to problems of moral reasoning, in our surviving texts neither the likelihood of option attainment nor the intensity of desire is quantified in any very precise way. The potential value for *us*, today, of tracing both differences and similarities between the ancient and modern accounts of how value is calculated by instrumentally rational agents is among the primary justifications I can offer for this book.¹⁷

In the course of the next several chapters, I will attempt to demonstrate that there actually was an “ancient Greek folk theory” of practical reasoning that centered on instrumental rationality. That is to say, the internally coherent set of assumptions about interest-based choice and action sketched above was widely known in Greek antiquity. Moreover, it was widely adopted as a *point d'appui* – although not necessarily as a satisfactory account of either observed behavior or morally choice-worthy action – by Greek writers who concerned themselves with ethics, politics, and history. They employed the theory as a foundation for causal explanations and evaluative assessments of human behavior and of social and political development.

The theory was adapted and refined over time. Few if any of the thinkers with whom we will be concerned supposed that the folk theory explained all relevant human behavior or all aspects of development. It failed, for example, to account for human aspirations to identify objectively correct ends; to act well, finely, and justly; to “do the right thing, in the right way, for the right reasons.” Understanding why people, individually and in groups, act as they do in different contexts required substantial adjustments to the folk theory’s premises. But, as I will hope to show, the base-line theory to which the adjustments were made was taken as a valuable starting point by Greek thinkers. The folk theory sets the frame for the study of Greek thought about the rationality of choice, by individuals, groups, and states; about behavior that conforms to predictions arising from the theory; and about behavior that deviates, in various ways, from those predictions.

The Greek folk theory of practical reasoning held that under ordinary social conditions (as opposed to thought experiments in which constraints on an agent’s choice are removed), the choices made by an advantage-maximizing agent are constrained by choices made (or that might be made) by others. The choices made

by oneself and others are in turn constrained by choices made in the past and canonized as institutions and culture: by laws, norms, shared beliefs about justice and morality, and by habitual behavior developed in the context of the rules established through those laws, norms, etc. The ubiquity and complexity of constraints on any given individual's choices make it very difficult to infer an agent's actual preferences and beliefs from his actions.

In order to capture the underpinnings of self-interested decision-making "in the wild," it was necessary to imaginatively strip away all of those constraints. That is the daunting imaginative task that Socrates' interlocutors undertake in the second book of Plato's *Republic*. That Plato's goal was to contrast the choices that would be made by an imaginary unconstrained, rationally self-interested agent with the choices made by an individual constrained by the choices made or potentially made by others becomes clear when we compare Plato's tale of "the ring of Gyges" with the Gyges story presented by the historian Herodotus.

1.2. The chosen path: Glaucon's thought experiment

By the end of the first book of Plato's *Republic*, a conversation on the topic of justice has reached the dead-end (*aporia*: literally "no-path") characteristic of Plato's earlier Socratic dialogues. The first book concludes with Socrates' refutation of the sophist Thrasymachus, who had argued for one version of what we are calling the folk theory of practical reasoning.¹⁸ Thrasymachus bluntly rejected the arguments made by "conventionalist" proponents of the folk theory, those who sought to show that it was actually in the best interest of every individual to act justly, according to established norms, and to obey the laws.¹⁹ Thrasymachus asserted that justice was nothing other than the advantage of the strong. The strong, those who know what they want and who are capable of getting it, should satisfy their desires by taking whatever they wish from the weak, and strong rational individuals will do just that. In political terms, this meant that the strong rule as tyrants, thereby using political power to maximize their own happiness at the expense of others (*Republic* 344d).

Socrates' questioning yielded the important point, asserted by Thrasymachus and not contested by Socrates, that the true "craftsman" (*demiourgos*) of self-interest never errs in his assessment of how best to pursue his own interest: "A ruler, insofar as he is a ruler, never makes errors and unerringly decrees what is best for himself" (340e-341a).²⁰ Thrasymachus' ideal craftsman of self interest resembles the choice theorist's idealized rational agent, who employs full information and unlimited cognitive capacity when calculating how his highest-ranked preferences may best be satisfied.

Socrates ultimately leads a blushing Thrasymachus to concede, grudgingly, that justice is virtue and wisdom (350d) and that injustice is never more profitable

than justice (354a).²¹ But Thrasymachus clearly remains unconvinced by the propositions to which he has been led to accede, and Plato's readers are still in the dark about why, if justice does not reliably produce the ordinary goods that most people seem to want, it is desirable as an end. In the second book of the dialogue, two of Socrates' interlocutors, Plato's brothers Glaucon and Adeimantus, propose a new beginning for the inquiry. Claiming that Thrasymachus was charmed by Socrates, "too soon, like a snake,"²² Glaucon proposes to revive Thrasymachus' argument, predicating it directly on the beliefs of ordinary people about how choices are made.

Glaucon announces that he will first state what kind of thing "people say justice is and where it comes from." Then he will show that those who practice justice do so unwillingly, as necessary rather than as good, for the life of an unjust person is much better than that of a just one, "as they say." This is not, he hastens to add, his own opinion, but an argument he has heard from Thrasymachus "and a myriad of others" (358b-c).²³ In brief, the two brothers plan to produce a philosophically rigorous version of the Thrasymachean version of what I have been calling the Greek folk theory of practical reasoning. They evidently believe that the standard "conventionalist folk-theoretic" arguments for the rationality of legal obedience and behaving in accordance with justice can be defeated by improved Thrasymachean arguments. Thus, if justice is to be defended, it must be on a new and firmer foundation.²⁴

The brothers' explicit goal is to prepare the ground so that Socrates can proceed to demonstrate that justice ought to be pursued as a rationally choice-worthy end in itself, even after it has been strictly divorced from all extraneous instrumental considerations – that is when acting justly is good for nothing other than justice itself. In order to embark on that enterprise (which will not be completed until the last book of the long dialogue and requires the construction of a theory of moral psychology and an ideal state), Plato's readers must have a prior understanding of what most Greeks considered to be human motivation and choice. We need to grasp what lies behind an ostensibly rational, but non-philosophical, individual's choice to act in some specific way, when he or she is seeking to achieve some desired end or goal. Before coming to terms with the moral-philosophical definition of ideal rationality as an end, a definition that Socrates will develop in the course of the dialogue, readers must be reintroduced to an ordinary, but typically cryptic, sort of practical reasoning. This is human reason understood as instrumental rationality: the decision processes and strategic choices that, according to modern theories of rational choice, underpin goal-oriented human behavior in the everyday phenomenal world.²⁵

Like Socrates, Glaucon and Adeimantus believe that ordinary people are *not* rational in the extended, moralized, sense of being motivated by reason to seek to

learn what is truly best and, having identified it, consistently going for it by employing beliefs about the world that amount to knowledge rather than opinion: beliefs that are true and justified. They want Socrates to demonstrate to them, in a form that will survive attacks by critics, that, despite appearances to the contrary, justice (being a just person and always acting accordingly) is more desirable than the ends that people ordinarily seek. They want Socrates to show that justice is not reducible to an instrument, a matter of words and deeds aimed at any goal other than ultimate and genuinely highest ends. They want Socrates to refute the argument that a truly rational person would ever employ practical reasoning to gain any outcome in preference to justice. Glaucon and Adeimantus accordingly devise a philosophical challenge that is designed to be as demanding as possible, so that Socrates' final demonstration, if successful, can be defended against objections that the ultimate value of justice as an end was not tested against the hardest imaginable case.

The challenge takes the form of an extended thought experiment. Glaucon first establishes that, according to a commonly-held view ("they say..."), people originally established and now obey law only because no one is in a position to do injustice (seek only his own advantage) with impunity in the face of similarly self-interested others (see chapter 2). He then sets out the counterfactual condition to which Socrates' defense of the inherent value of justice must be a refutation:

We can see most clearly that those who practice [justice] do it unwillingly and because they lack the power to do injustice, if we construct something like this in thought (*dianoia*): we grant to an [ostensibly] "just" and an [explicitly] unjust person the freedom (*exousia*) to do whatever they like. Then we may follow along, seeing where desire (*epithumia*) leads each one of them. We should then catch the just man, self-revealed, going along to the same destination as the unjust man, because of the striving to gain more and more (*pleonexia*) that every creature by its nature pursues as a good (Plato, *Republic* 359b-c).²⁶

The image that Plato invites his readers to form in their minds is one of following and observing someone as he travels on a chosen path, aiming to get to his desired destination. The translation of G.M.A. Grube (revised by C.D.C. Reeve: Hackett edition, 1997) makes this explicit: "we'll catch the just person red-handed travelling the same road as the unjust."²⁷ Glaucon explains that the reason that both persons will, if free to do so, follow the same path is the universal and natural drive of *pleonexia*: striving to get more and more.²⁸ Fulfilling that general preference to get more of whatever one happens to desire, to maximize advantage (here defined, with Thrasymachus, egoistically – as advantage *only* to oneself), he concludes, is

what any individual would naturally choose as good for himself in a condition of pristine freedom in which his highest-order desires aligned neatly with his expectations.

The natural, unconstrained choice is, Glaucon claims, not often in evidence in the phenomenal world because it is deformed whenever people are forced by social convention (*nomos*) to respect issues of equity (*ton ison*: 359c). The difference between the apparently just and the explicitly unjust man, Glaucon suggests, has nothing to do with their preferences over outcomes. It is entirely a matter of the reputation each happens to have. The “just man” is not truly just, in that he acts out of moral principles different from those of the unjust man. He is only apparently just: He has a reputation for being just that he does not deserve. Either he has not been caught breaking rules or he is too afraid of the likely consequences to take the chance of doing so. The “unjust man” acts in accordance with his preferences and is unlucky enough to have been discovered doing so. He thereby gained a bad reputation for breaking the same rules that the “just” man would break if he could do so with impunity. Glaucon here asserts that the ultimate driver of human action, the motivation of all freely chosen behavior, is the attempt to satisfy personal, subjective preferences.

The deeper truth about human motivation is assumed, in Glaucon’s challenge, to be obscured by what we might call the noise or, perhaps better, adopting Grube’s road metaphor, the “heavy traffic” of *ordinary human behavior*. That is: behavior under conditions in which each individual’s choices are constrained by considerations arising from the responses of others to his or her choices. The relevant considerations can be summed up under rubrics of rules, norms, laws, reputation, and piety. The basic idea is that, because the road to our most-preferred outcome is so often blocked by social traffic, we must ordinarily take less-desirable paths and end up at less-preferred destinations. But, freed from these constraints, absent any concern for the *social consequences* of acting according to our actual preferences, each individual would ignore inconvenient rules, and would always travel the road that leads directly to the expectation of having more and more of whatever is most desired. It is the need for a thought experiment that is capable of clearing the road of the traffic of ordinary behavior, in order to test intuitions about justice and choice, that is the context in which Glaucon tells the seemingly fanciful story of a shepherd who became a king – the first extended tale in a dialogue famous for its memorable images, allegories, and stories – the Line, the Sun, the Cave, the Three-headed Beast, and so on.

1.3. The story of Gyges’ ring

Glaucon’s story concerns “an ancestor of Gyges the Lydian.”²⁹ Following a convention long-standing among commentators on this passage, I will simply call

him Gyges.³⁰ At the beginning of the story, Gyges is a shepherd in the King's service – and so, presumably the reader is to imagine (based on Greek social conventions) that he was poor, lacked a social network that included influential persons, and spent much of his time alone with his flock. One day a thunderstorm and earthquake opened a chasm at the place where Gyges was tending his sheep. Gyges descends and sees many wonders. The only wonder that is described is a hollow bronze horse. It has windows in its sides, through which Gyges spots an over-sized corpse wearing only a gold ring. From this point in the story to its end, Gyges acts completely selfishly. Gyges takes the ring and soon discovers that it has the power to render him invisible. Gyges then arranges – how, we are not told – to become “one of the messengers sent to report to the King.” After he arrived at the palace, he “seduced the King's wife, attacked the King with her help, and took over the kingdom” (359c-360c). So, the story ends with Gyges, having ascended from the bottom to the peak of Lydian society, apparently amply rewarded for acting entirely in his own self-interest.

Having completed his tale, Glaucon stipulates that the two hypothetical persons, one thought to be just and the other unjust, are provided with invisibility rings.³¹ He specifies that the unjust man, for his part, will do what he pleases in complete secrecy. Moreover, taking over a key concept from Socrates' earlier conversation with Thrasymachus, Glaucon specifies that he will be a perfect craftsman of self-interest (360e-361a). Either he makes no errors (that is, he will always make choices that fulfill his preferences for having more and more) or, failing that, he will always be able to correct his errors. Glaucon proposes that under these conditions, the perfectly unjust and the ostensibly just man, as noted above, would behave in an identical self-serving way: He lists as the predicted behaviors, stealing goods from the market with impunity, entering other men's homes to have sex with whomever he pleased, and doing all the other things that would make him equal to a god among humans.³² Both men would, he concludes, proceed to just the same place (360c).³³

Unsurprisingly, these predicated behaviors track the actions of Thrasymachus' unjust craftsman of self-interest and of Gyges. Plato's Gyges story begins with the theft of the ring under presumed conditions of secrecy: Gyges is evidently alone when he enters the chasm, and even if someone happened upon the chasm later, no one need know that the corpse had been wearing a ring, or that Gyges was ever there. Gyges then uses the ring to enter the palace, have sex with the queen and, having killed the king with her help, to gain political power – the control of the kingdom.³⁴ Thus, we are to suppose that the basic goods that would be pursued by any person unconstrained by social rules reduce to more possessions, greater access to sex, and superior political power.

The value of the ring of invisibility to the thought experiment is underlined when Glaucon specifies that the rule-breaking of the unjust man will *never* be found out – so, not only will there be no punishment, but there will be no reputation cost, now or ever, for doing exactly as he pleases (361a). Moreover, Adeimantus adds a lengthy addendum to the thought experiment, specifying that there will be no divine retribution for rule breaking. Despite all his misdeeds, the perfectly unjust man will be loved, not hated, by the gods, because with his stolen wealth he will be able to perform all requisite sacrifices and other god-pleasing rituals. So, the unjust individual is unconstrained by “third party” divine authority or concerns about punishment for wrong-doing in some putative afterlife. Gyges, we may suppose, really will live happily ever after – at least according to the egoistical standard of having more and more.³⁵

Finally, neither in the original Gyges story, nor in the follow-up account of the two (“just” and unjust) men, is there any meaningful element of chance (at least after the ring has been found and its powers discovered). Gyges and the “two men” in Glaucon’s thought experiment act as if they were certain of the outcomes of their choices and actions. The absence of uncertainty, chance, or risk, in the experiment is underlined by Adeimantus’ addendum stipulating the permanence of good reputation and the benevolent attitude of the gods.

1.4. Plato on rationality of means and ends

The conclusion of the thought experiment, then, is that every individual, when completely free from extraneous social considerations, will seek to satisfy his own preferences, pursuing his own self-interest in a purposeful manner, and that he will identify his own self-interest as gaining possession of desirable goods. A person who achieves that position will be “equal to a god.” The goods in question are summed up, in the Gyges story and in Glaucon’s summary, as consumer goods (the things that can be bought or stolen in the market), access to sex, and power over others. Because the unconstrained individual expects that such things will make him maximally happy (i.e. godlike), we may call this the premise of *maximization* of expected subjective value, happiness, or advantage. Thus: Premise 1: expected value is maximized by making choices that, given the agent’s beliefs about the state of the world, will result in satisfying his highest-ranked preferences.

A second, implicit, premise complements the first. Premise 2 stipulates that an individual’s preferences over outcomes are rank-ordered: (1) having wealth/sex/power; (2) living according to the norms of justice/law; (3) being punished for lawbreaking. While we do not know how Gyges or the unjust man ranks wealth against sex or power, Gyges demonstrates unambiguously by his actions upon the opening of the chasm, and then upon learning the power of the ring, that he prefers the package of wealth, sex, and power to acting according to a

conventional conception of justice and not having those goods. After all, he could have chosen to leave the ring on the finger of the corpse, could have continued his life as a shepherd rather than arranging to become a messenger, could have left the King's wife alone and not attacked the King's person. But, clearly he expected to gain greater advantage by stealing, seducing, attacking, and ruling.

Likewise, Gyges' apparently rule-abiding life before gaining the ring (he was a shepherd not an outlaw) shows that (like most people under ordinary circumstances of "heavy traffic") he preferred *not* to suffer the likely consequences of being caught breaking the established rules: In conformity with Glaucon's description of the everyday behavior of most people, before the chasm opened Gyges presumably behaved as if he respected equity. But he did so only because of the external sanction of the social rules that kept him in his humble station, and the corresponding likelihood of bad consequences to himself of breaking them. So, we may say that if X is having the preference-satisfying goods, Y is living an ordinary rule-obedient existence, and Z is a high likelihood of paying the penalty for being caught breaking the rules, Gyges' preferences are ranked in the order $X > Y > Z$. And these ordinal preferences are *transitive*: that is to say, he never prefers Z to X.³⁶

Plato makes a point of emphasizing that behavior that accords with the two premises is by definition *rational* behavior: Glaucon states that, for someone who (like Gyges with the ring) *does* have the power to satisfy his preferences just as he pleases, to accept any limits on that power by making a binding agreement with someone else to abstain from doing injustice would be seen as irrational in multiple senses. He would be regarded as "insane" (*mainesthai*: 359b). Moreover, such a person would be thought of as "completely wretched and utterly senseless"³⁷ by any honest observer of his situation (i.e. anyone not constrained to say otherwise by worries about social rules: 360d). Plato links the conception of rationality as correctly calculating the means to one's most desired ends with an assumption that those ends will be egoistically self-interested by, as we have seen, having Glaucon borrow Thrasymachus' stipulation that the unjust man must be a true craftsman (*demiourgos*) of self-interest, unerring in his calculations of what is best for himself. The perfectly unjust man will never make mistakes in regard to what is obtainable and what is not (360e-361b).³⁸

The two premises, concerning transitivity of ranked preferences and actions calculated to maximize expected advantage, roughly track the definition of rationality in contemporary choice theory.³⁹ For Glaucon's rational exemplar, expected subjective value replaces what classical philosophers, including Plato, took as the correct objective standard – *eudaimonia* – as the condition that is maximized by the rational individual. This matters, for our purposes, insofar as *eudaimonia* was understood by the classical philosophers to concern *true* human interests. For Plato (along with Aristotle and other ancient moral philosophers) *eudaimonia* was an

objective state. The theory of rationality that Socrates will develop in the course of the *Republic* was normative in a moralized sense that theories of instrumental rationality, ancient and modern, are not: Socrates' establishment of the two premises of rationality concern not merely preferences and beliefs, but *right* preferences (that is for ends that are *genuine* goods) and *true* beliefs about the value of outcomes as well as about how desired outcomes are best obtained.

In order to maximize (moralized, objective) *eudaimonia*, the choice-maker's actions must follow from the expectations that emerge when both preferences and beliefs are morally, objectively correct. Since (per Plato's metaphysics) seeming and being are radically disjoint realms in ordinary human affairs, quotidian preferences and beliefs typically come apart from right preferences and true beliefs. They are only securely and consistently conjoined under the very special conditions pertaining, for example, in the *Republic's* ideal state of Callipolis. There, a fully-informed rule-breaking Gyges would in fact prefer Z (punishment) to X (unjustly taking sex, wealth, power) because he will realize that, having done wrong, punishment (in the form of correction of his mistake in pursuing an end that is not objectively valuable) is in his true interest: It will maximize his true advantage. We will return to *eudaimonia* in chapter 8. For our present purposes, the important point is that in Book 2, Plato has designed a thought experiment aimed at clarifying how most people in his original Greek audience would make choices and act in regard to others if they could somehow manage to get free of social "traffic." The thought experiment clearly establishes the grounding of rationality in preferences and beliefs, and draws a causal arrow pointing from the expectations associated with sets of preferences and beliefs to the choices and actions made by individuals.

The story of Gyges' ring advances the instrumental rationality thought experiment by, in contemporary parlance, backing away from the complexities of a game with multiple players, to a decision that centers on the psychological processes involved with individual choice. The introduction of the conceit of the ring of invisibility allows Gyges to satisfy his preferences without constraint, to behave in unproblematically (within the realm of subjective preferences and beliefs) advantage-maximizing ways, with complete confidence in the outcomes that will follow from his choices and actions. Because his road is traffic-free, he can expect that his highest-ranked preferences will be consistently achieved through his actions; his choices will result in his most-favored outcomes. Thus, ironically (in that it is achieved via invisibility), Gyges' behavior may be redescribed as directly reflecting his *revealed* preferences. The relationship between Gyges' observable (to the reader of the story, although not to the other characters in the story) choices and his private preferences over outcomes is disambiguated: Gyges is freed from the noisy considerations that arise with the responses of others to his actions. No

consideration for social rules impedes his choice; no other person acts in ways that would require rational Gyges to detour from his most-desired path.⁴⁰

Here we can see the utility of Plato's Gyges story for clarifying the groundwork of a theory of ordinary practical reasoning: It solves the problem, which is basic to rational-choice economic theory, of how to determine a given choice-maker's preferences. The normal approach to the revelation of preferences in the literature is based on market behavior: Any given individual J, with limited resources (budget B) has a choice between two different, same-priced indivisible goods X and Y (we assume $X \leq B$ and $Y \leq B$, but $(X + Y) > B$; so J cannot have both). J's choice of X reveals J's ordinal ranking of preferences: $X \geq Y$ (note that the choice might have been a mere coin-flip, if J were indifferent between X and Y). If the price of X is higher than the price of Y by (quantity) q, and if J still chooses X over Y, then we can begin to measure *by how much* J prefers X to Y: $X \geq (Y + q)$.⁴¹

The problem with preferences that are revealed by J's purchase of some good in a market is that the choice of X over Y (say, for example, X = flowers over Y = a bottle of whisky) reveals *only* that some specific J chose X over Y under the specific conditions pertaining at the moment of the purchase. While the goal of the theory of revealed preferences is to determine *consistent* preferences (at least for one individual choice-maker) over time, it is perfectly possible that under different conditions J will choose Y. Say that on an ordinary day, J will choose to purchase whiskey rather than flowers: His revealed preference is whiskey. But suppose we observe J on the day of his wedding anniversary: He chooses flowers over whiskey because he dreads facing the disappointment of his spouse, or the disdain of society. Under the circumstances, he ranks relationship and reputation over whiskey and chooses flowers accordingly. But were it not his anniversary, he would buy the whiskey.⁴²

Glaucon's thought experiment aims at revealing the "primitive" (in the sense of basic rather than vulgar) preferences that, he posits, lurk behind the readily observable, via agents' choices, constrained preferences of everyday life. He suggests, at a minimum, that every rational person seeks to maximize his own expected advantage, and maximizing advantage means that X: having more of a set of subjectively-valued goods, will be consistently chosen over Y: not having or having less of those goods, which will in turn be chosen if there is a high enough risk of Z: suffering the consequences of being detected in violating social rules and punished accordingly.⁴³

But Plato's Glaucon is even more ambitious: He suggests not only that Gyges' unconstrained preferences are complete, transitive, and stable over time, but also that (unless and until Socrates can demonstrate the value of justice as an inherently choiceworthy end in itself), *all* individuals would make similar choices under similarly "traffic free" conditions. Everyone (at least every adult male, "real man":

359b) is assumed to have similarly ranked primitive preferences over outcomes, reducible to a general preference simultaneously to have superabundant wealth, sex, and power (359c). This generalization seems precarious, at best, insofar as it does not accurately describe Socrates' preferences, nor presumably those of Glaucon himself. But even if Glaucon's posited set of egoistic primitive preferences cannot be universalized, it may nonetheless be the case that some or even many are motivated by them.⁴⁴

In the Greek political tradition, and as Plato's Thrasymachus pointed out (344d), these wants are strongly associated with the figure of the tyrant. Gyges suggests, therefore, that an "ordinary man" – even a poor shepherd – would act as a tyrant if he were free from social constraints. It is essential to note, however, that the basic "folk theoretic" logic of rational choice, as sketched by Xenophon's Socrates in the *Memorabilia* passage (based on preferences and beliefs), need *not* imply that interests (outcomes sought) are egoistical (self-interested in the narrow sense of discounting the interests of others), much less that interests can be reduced to the primitive triad advanced by Glaucon in *Republic* book 2. The distinction between the core logic of rational choice and the particular ends that might be pursued by a rational agent means, in turn, that the fundamental "folk theoretic" architecture of instrumental rationality remains available to Plato (as to other eudaimonists), once it has been separated from "Thrasymachean" assumptions about the kind of ends that would be sought by a rational agent in an unconstrained condition.

In book 8 and 9 of Plato's *Republic* the political implications of instrumental rationality, uncoupled from the right preferences and true beliefs that have been specified in the intervening books, are made manifest: The collapse of harmonious Callipolis is followed by a degenerative cycle of regimes that leads to the diversity of individual desires, absence of legal constraint, and consequent individual freedom of choice that Plato associates with democracy (see chapter 5). With the ultimate degeneration of democracy into the rule of the tyrant, political order is reduced to the expression of the primitive preferences specified by Thrasymachus and Glaucon. Plato's point here is to reveal the profound personal unhappiness that attends that expression. While seemingly instrumentally rational according to the Thrasymachean variant of the folk theory, the tyrant's choices lead him to the depths of misery. As will be evident in the following chapters, exploration of the implications of equating unconstrained egoistical self-interest with tyrannical desires and behavior is not unique to Plato and the characters in his dialogues.

A primary conclusion of Plato's *Republic*, in books 9 and 10, is that the tyrant is utterly miserable, his apparent freedom of choice leaves him as an object "slave to slaves."⁴⁵ And so a truly rational individual (one who is rational in the moralized eudaimonistic sense of consistently choosing objectively valuable ends, not merely in choosing the best means to subjectively valued ends) would never step onto the

path leading to that destination. But, meanwhile, back in book 2, Glaucon has established the preconditions of a theory of practical reasoning that is free from the complicating factor of primitive preference diversity.

Glaucon's approach here might be likened to a "vulgar economist" approach to choice, one that assumes that all preferences are reducible to material welfare and so can readily be expressed in dollars (or some other currency). Contemporary choice theory typically assumes, on the contrary, that different people want quite different things; rationality is not predicated on *what* people prefer (rationality over ends), but on the completeness and transitivity of ranking among whatever preferences they happen to have, their beliefs about the state of the world, and the alignment of their actions with the resulting expectations. Thus, when a choice theorist sets a quantity (i.e. a "utility function") on a player's preferences, the quantity represents what is assumed to matter (based on revealed preferences and beliefs) to that player, not to everyone.⁴⁶ As we will see in the coming chapters, the Greek tradition of thinking about the rationality of choice offered some thoughtful alternatives to Glaucon's reductionism concerning choice-worthy ends.

1.5. Gyges' ring decision tree

We are now in a position to illustrate Plato's Gyges story, employing the standard "tree form" used in contemporary decision theory. See Figure 1.1. The extensive form may be understood as a way to give a radically simplified and concrete visual expression to Plato's story, by employing the image, evoked by Glaucon in the opening quotation, of the choices made by a rational person travelling a chosen path. The main road (the path that is predicted to be taken by the rational choice-maker) is indicated by the long dark line. Potential alternatives to the main road are indicated by lighter, shorter lines. At each intersection, or "node," a choice is made by a decision-maker. Here, the decision tree illustrates a parametric choice (i.e. an environment which the relevant parameters are fixed) rather than a non-parametric strategic game, because the only "player," other than Gyges (G) himself, is "Nature" (N).⁴⁷ Unlike other players in strategic games, Nature has no preferences over outcomes and does not act strategically.

[Figure 1.1. Gyges decision tree about here]

The outcomes for players with preferences are indicated at the end of each path. Here the outcomes are listed as α to ζ . The value of the outcome for each player (here, only Gyges) is specified by a payoff quantity, in this case set at 0-4 and meant only to capture Gyges' ordinal ranking of preferences over outcomes (lowest to highest) rather than the cardinal value of each outcome (how much weight he attaches to each outcome relative to each other outcome). Listing payoff quantities

thus allows us to indicate the ranking of the preferences over outcomes of the player, who, per above, is assumed to be a rational expected advantage maximizer, with complete, transitively ordered preferences. When it comes to a choice between two or more outcomes, the player will always choose the available outcome with the highest payoff.

In the “Gyges decision tree” (Figure 1.1) we begin, as usual, at the “root” of the tree; in this illustration the root is in the lower left corner of the figure. “Nature” (N) moves first, either revealing or not revealing the ring to Gyges. N is added to this game to indicate that there is an element of chance involved. Since Plato does not suggest that the ring was revealed through the will of some purposeful divine authority, we can suppose that the chasm opened by accident. It might not have opened when and where it did, and so Gyges might have lived out his life as a shepherd. If the ring were never revealed, then we would end up at outcome ζ ; Gyges would get a payoff of 0 and the process would end there.

But, as we know from the story, N did reveal the ring; since we suppose that N has no actual volition, it just so happens that the ring is revealed. Since it is revealed, we proceed to the second node (fork) on the path; here Gyges (G) moves, choosing either to ignore the ring and remain a shepherd, or to take the ring. At the third node he chooses whether to become a messenger or remain a shepherd. At fourth node he chooses whether or not to seduce the Queen. At the fifth and final node he chooses between killing the king, marrying the queen, and ruling the kingdom and not doing these things.

We can determine what choices a player makes in the game by a process of backwards induction – even if we cannot say whether this form of reasoning was used by the ancient Greeks. In this case, after N has revealed the ring, Gyges is the only choice-maker in this decision matrix, meaning that Gyges does not need to take into account the choices of any other actors. But as in some strategic games, we will assume that the choice-maker has full information (he is a “perfect craftsman,” per above): That is, Gyges can look ahead (up the decision tree) to the choices over outcomes that he will face and he can calculate the payoff associated with each outcome. We begin the process of backwards induction at the last (fifth) node, in upper right hand corner of the figure. Gyges chooses α (payoff 4: wealth, sex, and political power) over β (payoff 3: having seduced the queen he has got some wealth and sex). So, thinking of the diagram as a tree, we can cut off the β branch (not killing the king), since it will never be chosen.

Next, backing up to the fourth node, Gyges now chooses between the surviving outcome α and γ , which has a payoff of 2 (being a messenger is more lucrative than being a shepherd). Since $4 > 2$ Gyges chooses to seduce the queen, and we cut off branch γ (not seduce). At the third and second nodes, Gyges chooses between α and δ (payoff 1, since he has the gold ring), and thus becomes a

messenger, and then between α and ε (payoff 0, he has not improved his situation in any way) and so takes the ring. In each case, the payoff for α (4) exceeds the alternative outcomes (1,0), so δ and ε are cut off. With all the branches (alternate paths and outcomes) now cut from Gyges' decision tree, his path forward from the second node (his first choice) to the outcome α is obvious: he will take ring, become the King's messenger, seduce queen, kill the King, marry the Queen, and rule the kingdom.

1.6. Why choice theory in Greek dress?

Obviously, Plato did not include an illustration of a decision tree in his original written text of the *Republic*. But Glaucon's reference to following the choices made by a rational person traveling on an imagined path to a chosen destination suggests that Plato meant the reader to visualize something like an extensive form decision tree when considering Gyges' choices and their implications for human motivation under ideal-type unconstrained conditions. *Mutatis mutandis*, the extensive form decision tree looks somewhat like, and has much the same function as, standard illustrations of the Line (*Republic* 509d-511e), which modern editors of Plato's *Republic* produce to clarify the complex metaphysical-epistemological argument of Book 6 (e.g. Hackett edition 1997, p. 1130). Moreover, in his discussion of the Line, Plato himself calls attention to his contemporaries' practice of drawing figures to illustrate concepts in "geometry, calculation, and the like" (510c-d). It does not, therefore, seem too much of a stretch to say that, in *Republic* book 2, Plato has sketched a theory of instrumental rationality and that he (and his intended readers) would quickly grasp the pedagogic value of decision trees and other illustrations.

While Plato did not employ shorthand notation, in the form of letters to indicate conditions, or the methods of formal logic, when discussing preferences and how they are rationally ordered, Plato's student, Aristotle, certainly did. In a passage of one of his fundamental works on deductive reasoning, the *Prior Analytics* (2.68a), Aristotle discusses preference ranking in a way that is strikingly reminiscent of contemporary decision theory: "When, of two opposite alternatives A and B, A is preferable (*haireteron*) to B, and similarly D is preferable to C, if A and C together are preferable to B and D together, A is preferable to D."⁴⁸ Aristotle goes on to clear up any confusion about the ranking, by describing A and B, D and C as respective polar opposites, and by specifying that A and D are goods to be pursued, while C and B are evils to be avoided. So we might illustrate the argument in one dimension, on the models of Plato's allusion to illustrations in geometry etc. and contemporary illustrations of Plato's analogy of the Line, as in Figure 2.

[Figure 1.2 about here]

If we take *hairteron* to be equivalent to “greater than” (more highly ranked in a preference ordering) Aristotle’s description of the preference ordering may be written as $A > B, D > C, (A+C) > (B+D) \rightarrow A > D, C > B$.⁴⁹

Aristotle’s example of preference ranking concerns the attitude of a lover towards the relations he may have with his beloved:

If then every lover under the influence of his love would prefer his beloved to be disposed to gratify him (*charizesthai*) (A) without actually doing so (C), rather than actually gratifying him (D) without being inclined to do so (B), clearly A – that the beloved should be so inclined – is preferable to the act of gratification. Therefore in love to have one’s affection returned (*phileisthai*) is preferable to having sex (*sunousia*) with the beloved. Therefore love aims at affection (*philia*) rather than at having sex; and if affection is the principle aim of love, it is also the end (*telos*) of love. Therefore having sex is either not an end at all, or only with a view to receiving affection.⁵⁰

Aristotle thus suggests that “every lover” will have a complete and transitive preference ordering, ranked as follows:

- A Affection (*philia*), having the beloved’s love and “disposition to gratify” (even without sex)
- D Having consensual sex (*sunousia*) with the beloved
- C Not having consensual sex with the beloved
- B No affection, not having the beloved’s love and “desire to please” (even with sex).⁵¹

He concludes this passage by asserting the generality of the result: “The same principle, indeed governs all other desires (*epithumiai*) and crafts (*technai*).” The key point, for our purposes, is that Aristotle’s approach to establishing a complete ordinal ranking of preferences as the basis for assessing the choices that a rational individual will make, appears rather similar to the approach used in contemporary theoretic approaches to rational choice and decision-making. While Aristotle’s work on formal logic was ground-breaking, there is no reason to think that Plato, or other Greek writers concerned with problems of choice, would have had any difficulty in following the argument, or (leaving aside the specific example of love and sex – which obviously contradicts Glaucon’s specification of primitive human preferences) any objection to the general applicability of his method.

I will hope to show, in the course of this book, that the approach of reading ancient Greek texts and institutions through the lens of elementary (mostly non-

algebraic) contemporary decision and game theory, and the toy models that contemporary theorists employ, sheds new light on the meaning of otherwise obscure passages and seemingly peculiar rules. It can illuminate certain relationships between texts. And it can clarify some of the working assumptions shared by the ancients who wrote Greek texts, by those who first read those texts, and by those who later critically discussed and built upon them. But even for those who accept that the approach sketched here is a potentially useful addition to the classicist's interpretive toolbox, it is in no sense a panacea.

Turning passages of artful and subtle writing into formal games is reductive – it strips away much that every reader of Greek texts (including myself) cares about. This reductive approach can only ever be an augment to the text: Likewise, studying an editor's illustration of Plato's Line cannot replace a careful reading of *Republic* Book 6. But, just as Plato's Glaucon sought to eliminate the noise introduced into real-world decisions by employing a thought experiment in order to better perceive basic motivations of ordinary people, reducing complexity can increase analytic clarity. Gaining greater clarity can in turn help us to better understand the complex original – whether it is a normative claim, a phenomenon of social interaction, an institution, or a passage in a Greek literary text or document. In brief, I will suggest that the “folk theory of practical reasoning” – an understanding of instrumental rationality as expected advantage maximization based on orderly preferences and coherent beliefs – underpins a good deal of classical thought. Excavating those conceptual foundations should enable us to be better readers of classical authors, better interpreters of classical history, and perhaps better students of human behavior.

It would, however, be reductive in the *wrong* way to suppose that instrumental rationality was the sum total, or even the primary concern of Greek writers concerned with the choices and actions of individuals and collective agents. As we will see, it was often the *limits* upon and *violations* of the sort of rationality exemplified by Plato's thought experiments in *Republic* book 2 that most interested Greek writers who concerned themselves with the problems of individual and social choice. They recorded and commented on the sometimes predictable and sometimes unpredictable mistakes made by choice-makers who imagined themselves as acting in a fully rational manner. And they were deeply interested in the behavior of those individuals and groups who understood their own motivations as something other than calculating self-interest.

Nevertheless, as I will try to show, it would also be wrong to conclude from these facts that ancient Greek writers were *not* deeply concerned with the folk theory, as sketched above. It is only against the background of assumptions about instrumental rationality that irrational or non-rational behavior becomes an object of analytic attention. Strategic choice-making is the point of departure for the large

body of contemporary empirical scholarship in behavioral economics and behavioral psychology. Empirical studies in those fields demonstrate that theories assuming perfectly rational choice-making fail to predict actual human behavior in real-world situations. But the object of such study is not to show that human behavior is simply irrational. Rather, it is to explore how, under what conditions, and with what regularity, real-world human choice-makers diverge from the predictions arising from choice theory. The situation is in some ways similar, or so I will argue, in ancient Greek political thought.⁵²

1.7. Herodotus' Gyges

If a theory of choice were limited to explaining the decisions made by an unconstrained individual, as illustrated by the "Gyges decision tree," its application to the social realm would be limited. But decision theory can include constraints (see chapter 6) and strategic game theory concerns two or more choice-making players. The complete freedom always to choose an outcome that reliably and fully satisfies personal preferences, as enjoyed by Plato's Gyges due to his power of invisibility, is unavailable to players in strategic games. Social constraint (by which I mean the choices of other agents, including those who enforce social and legal rules) and chance, the "traffic" that Plato sought to strip away with his thought experiment, are precisely what strategic game theory is concerned with. The point of the exercise is to demonstrate the choices that an advantage-maximizing player will make under conditions in which the desires and beliefs of another advantage maximizing player, with similar or different preference orderings, prevent the first player from achieving his or her highest-ranked outcome. Moreover, in some 2-player games, the situation of frustration is mutual in that *neither* player will get his or her preferred outcome. Under these conditions, to be considered in more detail in later chapters, the players are confronted with a dilemma: if each pursues his own interest the total subjective value available to them is, both jointly and severally, lower than the value that they could each enjoy if they could somehow find a way to cooperate without opening themselves to a sucker's payoff.

We can readily turn the 1-player Gyges' ring decision tree into a multiple (2- or 3-player) strategic game by considering the historian Herodotus' earlier, alternative story of Gyges' ascent to power in Lydia. Considering the likely intertextual relationship between the Gyges stories related in Plato's *Republic* book 2 and Herodotus' *Histories* book 1 (1.8-14) strengthens the presumption that Plato was self-consciously constructing, in words, a 1-actor parametric decision tree by eliminating the other choice-making agents who take key roles in Herodotus' account. Herodotus' Gyges story, like Plato's, is explicitly about making choices – the Greek vocabulary of choice (*hairêsis*) is prominently featured. But, in contrast to Plato's story, Herodotus' Gyges narrative consistently emphasizes the constraints

upon Gyges as a choice-maker by employing the Greek vocabulary of necessity (*anangke*).⁵³

In Herodotus' story, there is no ring of invisibility and Gyges' motives, other than his desire to stay alive, remain obscure. They are obscure because, in contrast to Plato's Gyges, each choice made by Herodotus' Gyges is constrained by the behavior of others.⁵⁴ Herodotus' Gyges is the bodyguard and confidant of Kandaules, king of Sardis (capital city of Lydia). Herodotus informs us that, madly in love with his (nameless in this story) wife and concerned that his bodyguard does not believe his description of her unmatched beauty, Kandaules orders Gyges to hide behind the door of the royal bedroom, and to observe the Queen as she undressed for bed. Gyges objects to the plan, calling it "lawless" (*anomon*). Herodotus specifies, however, that Gyges protested because he feared that the plan would lead to a bad outcome for himself (Herodotus, *Histories* 1.9.1).⁵⁵ We are given no reason to think that Gyges regards the badness of the outcome as the wrong inherent in violating a norm. Rather, per the King's assurances (below), Gyges seems to worry that the Queen might spot him and, if she were offended and vengeful, might punish him.

Kandaules is adamant, however, and Gyges, as a servant of the King, is unable to avoid the choice of either obeying the order to spy on the Queen, or suffering some (unspecified, but presumably terrible) punishment.⁵⁶ Kandaules assures Gyges that he, the King, will manage things so that the Queen will remain ignorant of the fact that Gyges has seen her undressed.⁵⁷ But he does acknowledge that there is some risk involved, since he urges Gyges to exercise care in exiting the bedroom in order to avoid being seen.⁵⁸ In the event, the Queen does spot Gyges leaving her bedroom. She realizes what her husband had done and is deeply shamed by the willful violation of her modesty. But instead of raising an objection, she conceals the fact that Gyges has been spotted, and quietly plans her own revenge upon the King as the instigator of the spying plan.

The Queen gathers trusted guards, summons Gyges, and presents him with a choice between "two roads" (1.11.2).⁵⁹ If Gyges takes the first road, he will agree to kill the king according to the Queen's instructions. At the end of this road, she says, is marriage to herself and rulership over the Lydians. On the second road, Gyges refuses to commit to killing Kandaules. At the end of this short road Gyges is immediately executed by the Queen's guards. As he had when Kandaules first ordered him to spy on the Queen, Gyges seeks to avoid the binary choice forced upon him by a powerful player capable of taking his life (1.11.3).⁶⁰

We are not told why Gyges begs to be offered a third option (neither killing nor being killed), or whether he is sincere in requesting it. Herodotus' language does not point to moral qualms: He does not say that Gyges objected that killing Kandaules would be *wrong*. By analogy with Gyges' reported worries about his own safety when ordered by Kandaules to spy, we may suppose that high among his

reasons for preferring a third way is rational fear of the risk to himself: The murder plot might go wrong; the Queen might have him executed after he kills the King. In these readily-imaginable scenarios, the payoff at the end of the first road is bad for Gyges. But because, like Kandaules, the Queen is adamant and the threat of immediate execution is credible (her guards are right there), Gyges “saw that dire necessity was truly upon him,” and “chose (*haireetai*) to go on living” (1.11.4).⁶¹ As it turned out, the murder plot succeeded: Gyges stabbed the King to death in his sleep. The Queen fulfilled her part of the arrangement; Gyges married her and became king of Lydia. Whatever Herodotus’ Gyges’ original preferences (beyond simply staying alive), he had achieved the rewards of sex, wealth, and power that Plato’s explicitly selfish, unconstrained Gyges had aimed at and had achieved by ruthlessly pursuing his own preferred ends.

Even without the ring, the similarity of Herodotus’ story to Plato’s is obvious: In each tale Gyges (or his ancestor), a servant of the King, has sex with the Queen (via seduction or marriage), murders the King, and comes to rule the kingdom. There is a substantial literature seeking to untangle the relationship between these two texts. Some scholars have suggested that both stories derive from some now-lost and more detailed third source that, hypothetically, combines the key features of Plato’s and Herodotus accounts concerning Gyges. While certainty in such matters is impossible, Andrew Laird (2001) has made a strong case that there is no third source, and that Plato deliberately adopted and imaginatively reworked Herodotus’ original tale for his own philosophical purposes.

I have argued that one of Plato’s goals in book 2 of the *Republic* is to clarify a folk theory of practical reasoning by eliminating the “traffic” of ordinary social behavior, conditioned as it is by expectations concerning other people’s actions. In Plato’s retelling, Glaucon implies that “every Gyges,” that is to say every rational individual, from slave to monarch, has similarly selfish ends in view. Laird’s thesis, that Plato took the story directly from Herodotus and reworked it for his own purposes, gains further plausibility when we reduce Herodotus’ story to an extensive game form and compare it to the decision tree illustrated above.⁶²

1.8. Kandaules, Gyges, and the Queen, 2- and 3-player games

We are, I suppose, encouraged to think of Herodotus’ story in something like an extensive game form by attending to the Queen’s reference to Gyges’ necessity of choosing between “two roads.” The allusion to a choice-maker at a fork in the road anticipates a decision node of just the sort that is illustrated by the extensive game form.⁶³ The Queen’s metaphor of two roads recalls Glaucon’s image of the single path that would be followed by both an ostensibly just and an explicitly unjust possessor of an invisibility ring, suggesting that, at least in their Gyges tales, Plato and Herodotus were working with similar choice-theoretic toolkits. The key

difference, of course, is that in Herodotus' story Gyges' highest-order preferences (other than preferring life to death) are not immediately revealed by his behavior. Because the Queen is a player and may, depending on Gyges' choice, choose to execute him on the spot, we can only guess at his preference ranking between "kill the king with some chance (depending on the Queen's honesty) of marrying the queen becoming king," and "neither kill nor be killed."

We know from Gyges' ultimate choice that he prefers to kill (and take his chances) than be killed. But we remain in the dark as to whether his request for a third way was (1) motivated by a preference not to kill the king, or (2) reflects his belief that the Queen was unlikely to fulfill her promise, or (3) was a tactic for gaining more information about her intentions. This lack of certainty is reflected in the limited range of Gyges' payoffs (0-1) in the games illustrated below. Plato's story, by revealing Gyges complete and transitive preferences through his unconstrained choices, eliminates the guess-work.

The fork described by the Queen is illustrated by the simple game tree in Figure 1.3. As in the Gyges' ring decision tree, quantitative payoffs are assigned to outcomes for each player (Queen: 0-2, Gyges: 0-1). As in the Gyges decision tree (Figure 1.1), the payoff quantities are meant to capture only the ordinal (not the cardinal) rankings of preferences over outcomes of each player.

[Figure 1.3 about here]

Gyges (G) moves first, choosing to kill the King with outcome α or not. If he chooses not to kill the King, then the Queen (Q) moves, choosing to kill Gyges with outcome β or not to kill him with outcome γ . The Queen is credibly committed to killing Gyges if he refuses her offer. Her commitment is under-written by comparing her bad payoff of 0 at γ (Q suffers the humiliation of having been spied upon, and the added burden of being disobeyed without consequences) to her relatively good payoff at β (1: she punishes the instrument of her humiliation). So, we cut off the branch leading to outcome γ . Gyges must, therefore choose between β and α – as the Queen succinctly points out, given her resolve, the road before Gyges has only two forks. Because Gyges' payoff for α (1: the payoff of getting the Queen and the kingdom is discounted by G's consideration of the risks of the plot failing, the Queen reneging, his reputation) is better than for β (0: he is killed) he chooses to accept the Queen's proposal. The predicted outcome of the game is therefore α : Gyges kills the King, hoping to marry the Queen.

The fork in the road alluded to by the Queen is the last node in a more complicated game played by Gyges, the Queen, and King Kandaules; the very simple game illustrated in Figure 1.3 can be thought of as a subgame of a complex 3-player (plus Nature) game. The extensive game form for Herodotus' tale of Gyges is more

elaborate than that of Plato's story because there are three, rather than just one, choice-making players, each with his or her own payoffs for possible outcomes. The full game is illustrated as a tree in Figure 1.4. As in the other games considered in this chapter, payoffs for each player indicate ordinal preference rankings for each player (King: 0-4, Gyges: 0-1, Queen: 0-3), from lowest to highest. Note that a given player's preference ranking may be the same for different outcomes. When this is the case, the player is assumed (for the purposes of the game) to be indifferent between outcomes (i.e. is willing to have the choice between them made by a coin flip).⁶⁴ Indifference between outcomes does not affect the result of this game, but can be highly relevant in other, more complex, games.

[Figure 1.4 about here]

In this game, King Kandaules (K) moves first, choosing to order Gyges to spy on his wife or not giving that order. If he does not order Gyges to spy, the game ends at θ with payoff to the King of 2 (he is denied the pleasure of knowing that Gyges has seen his wife naked). If K orders Gyges to spy, Gyges (G) chooses to obey or disobey. If he disobeys, K chooses whether to forgive or to kill him for disobedience. If G obeys, then Nature (N) determines whether or not the Queen (Q) spots Gyges spying on her. N also determines the Queen's "type" – that is, whether or not she is vengeful. Note that, because the players cannot, *ex ante*, know for sure how N will choose, there is a level of uncertainty in this game. If it were possible for the players correctly to measure that uncertainty as a statistical probability (e.g. 1:10 chance Q will spot G and is vengeful), we could assign quantitative weights to the players' revealed preferences (their choices), thus moving from simple ordinal ranking to more informative cardinal (intensity-weighted) preferences over outcomes.⁶⁵

If the Queen happened not to spot Gyges, or if, counterfactually, she happened to be indifferent to (or preferred) being spied upon, or if having been spied upon she just felt sad rather than angry, or if she was the type of person who expressed anger in ways other than planning vengeance, the game would end at ϵ , and there would be no bad consequences for Gyges or the King. If N determines that Q sees Gyges and is vengeful, Q chooses whether to kill Gyges outright, or to threaten G with death if he does not agree to kill the king. If Q threatens, we end up at the subgame illustrated in Figure 1.3.

As we already know from the backward induction that cut off outcomes β and γ , at the last node G will choose α . One node back, Q chooses between α (3: she gets revenge on Kandaules, against whom her anger is primarily directed) and δ (2: she punishes the instrument of her humiliation, but fails to punish the principal). Thus we cut off δ . As we have seen, N fails to determine outcome ϵ . Thus, one node back, G chooses between α (1, per above) and, ζ (0: he is killed by the King as

punishment for disobedience). His choice is limited to α or ζ because outcome η (1: G is forgiven by the King) is cut off by K's choice. Confronted with a disobedient Gyges, K will choose ζ (3: K has the satisfaction of punishing disobedience) over η (1: K's direct order is disobeyed, and disobedience goes unpunished, making future disobedience more likely). Thus G chooses α . The equilibrium path of the game is for G obediently to spy, then when N determines that Q spots G and is vengeful, for Q to threaten, and for G to agree to kill K, which results in his marrying the Queen, and ruling the kingdom.

1.9. Kandaules as imperfect craftsman of self-interest

Thus far in our backwards induction of this extensive form game, every player has chosen rationally, according to the ranked preference orderings listed in Figure 1.3. But at the very first (root) decision node, we encounter an apparent anomaly: with outcomes β through η cut from the decision tree, K chooses between α (0: K is murdered by Gyges) and θ (2: not giving the order to spy, he is denied some measure of pleasure, as above). As we know from the story, K rejects θ . Since $0 < 2$ (we assume death is worse than living without spying), the King has apparently acted irrationally: Why? Whether or not the Greeks used something like backwards induction in analyzing this kind of situation, the question of why Kandaules started down the path that ended in his death is surely a question raised by Herodotus' story.

Perhaps Kandaules is just clueless: i.e. not very good at the kind of calculation demanded by strategic games: thinking "up and back down the game tree" to map the expected consequences of his own and others' choices and to choose accordingly.⁶⁶ Perhaps he has a false belief about some relevant state of the world (e.g. that Gyges is so loyal that he would choose to be killed over killing his King). Perhaps, we have listed K's preferences incorrectly: Clearly, in Herodotus' story, Kandaules strongly desires Gyges to spy on his wife. But, although he is besotted with his wife, Herodotus does not depict Kandaules as preferring his spying plan to life itself.⁶⁷

Herodotus' story suggests, however, another explanation: Kandaules was rational, but either unlucky or poor calculator of risk. Along with the number of players, a difference between the Gyges with the ring game and the King, Queen, and Gyges game is the factor of uncertainty: At the Nature node, we introduce a set of factors that is not known with certainty: Will the Queen see Gyges? If she does see him, will she be vengeful and will her vengeance will be aimed at the King? Kandaules makes light of those possibilities: When encouraging Gyges to do his bidding, he expresses confidence in his own machinations, and downplays the likelihood that the Queen will spot Gyges. Kandaules appears to know that his wife *might* be watchful, since he urges Gyges to be careful in exiting the bedroom,

implying that if he is not sufficiently careful, he will be seen. The consequences of that possibility remain unspecified, but the King confidently assures Gyges that no harm will come to him from the Queen (1.9.1).⁶⁸ Kandaules seems to discount the possibility that a watchful and vengeful Queen might seek revenge on the King himself.

Greek writers on the rationality of choice were, as we will see (chapters 6, 7 and Appendix), very concerned with the concept of likelihood (*eikos*). Plato confidently asserted that it was possible for experts to correctly measure the probability of certain outcomes (*Republic* 467c-d; with discussion in chapter 6). As noted in the first section of this chapter, Herodotus did not express likelihood in quantitative terms, as mathematical frequency; nor did other Greek writers. Yet Herodotus' readers can make sense of this simple game even without resort to precisely quantified probabilities.

Herodotus' narrative suggests that Kandaules underestimated the likelihood, and thus the risk to himself, of the possible sequence "Gyges is discovered; the Queen is vengeful; she seeks revenge on the King; Gyges prefers life to loyalty." So perhaps Kandaules was just unlucky; he may have lost a fair bet with a high expected payoff and a small chance of a very bad outcome. But Herodotus' reader is prompted to conclude that Kandaules made a choice that would have been rejected by a more thoughtful individual, one who was better informed about his wife's and his bodyguard's preferences and beliefs. So, we may conclude that, unlike the choice-makers in Plato's *Republic* book 2 thought experiment, Kandaules was *not* a flawless craftsman of self-interest: He failed to factor relevant likelihoods into his expectation of the value to himself of Gyges' spying on the Queen.⁶⁹

We are now in a position to speculate on the source of Kandaules' error: Perhaps his judgment was corrupted by an unruly passion for his wife. Perhaps his royal position led him to overestimate his own cleverness. Herodotus' text offers support for each of those possibilities.⁷⁰ I would suggest that Herodotus meant the reader to think about each possibility (and perhaps others) – and to realize that it is not possible to come to a final decision about which was the decisive factor. If that is correct, we may say that Herodotus leads the reader to think about strategic choices in the face of uncertainty, but in a way that resists the notion that historically salient outcomes (like the death of a king) will be readily forecast by, or prevented through, the exercise of human reason. We, especially if we are readers not only of Herodotus' *Histories*, but also of Plato's *Republic*, are invited to consider the limits on the predictive power of what I have been calling the ancient folk theory of instrumental rationality.

1.10. Oracular information and choice

There is, as readers familiar with Herodotus' methods will anticipate, another possible answer to the question of why Kandaules erred, and one that is indeed offered in Herodotus' text: Kandaules' untimely death was necessitated by fate (1.8.2).⁷¹ So we might conclude that there is no value to thinking about instrumental rationality, because even a perfect craftsman of self-interest can never know when some divine force will upset all rational plans by putting a heavy thumb on the scales of the probable.⁷² But the conclusion that what is fated can never be a matter of accessible information, and thus that it has nothing to do with the rational choices made by humans, might be too hasty. Perhaps Kandaules was fated to come to a bad end only *because* of some combination of his unruly passion and his royal arrogance. Perhaps, even if fated, his bad end might have been forestalled (even if it could not be prevented) by more "craftsman-like" choice-making, utilizing the best available sources of information.

Herodotus *in propria persona*, like many of the characters in his *Histories*, was famously interested in oracles and he supposed that certain oracles, notably including the oracle at Delphi, could in fact reveal salient information relevant to human choices. So, the possibility arises that a true craftsman of rational self-interest would not make a highly consequential choice without gaining the information available from an oracle. The uncertainty represented by the N node might, if all went well, be removed or reduced by consulting an oracle. Of course, that requires, first, that the oracle be real (capable of providing relevant information) and, next, that the oracle's answer be either unambiguous or correctly interpreted by the player.

As every reader of Herodotus soon learns, those very demanding conditions often go unmet. But oracles, at least the sort of oracles described and accepted as valid by Herodotus and his characters in the *Histories*, make sense *only* in a world in which choice-makers assume that much human behavior can be understood as rational, in the "folk theory" sense sketched above. Herodotus introduces the story of Gyges by reference to his famous descendent and the last of his line, King Croesus (1.6). Croesus was interested in oracles specifically as instruments for gaining valuable information that could (he supposed) allow him to eliminate uncertainty when faced with a momentous decision. Herodotus' Croesus is in certain ways a model of the rational (although not mistake-free) oracle-user, first subjecting the oracle to what amounts to a blind test of its veracity, and then using its answers in making salient choices about foreign policy.

Plato ended the story of Gyges with murder and royal marriage, but Herodotus offers a coda: Gyges took the kingship and ruled *as a result of* the Delphic oracle (1.13.1).⁷³ The oracle was instrumental in consolidating Gyges' authority because it ended a civil war that broke out after Kandaules' fate was learned, between armed Lydian partisans of the family of Kandaules and the faction of Gyges.

Both sides eventually agreed (*sunebêsan*) that if the oracle at Delphi decreed that Gyges was to rule, he would do so by common agreement of the Lydians. If the oracle decided against him, he would return the throne to Kandaules' relatives. The oracle decided in favor of Gyges (1.13.2) who then sent many offerings to Delphi. These were the first such offerings from an Asian ruler; the gifts were consequential enough to be later called "the Gygian [treasure]" (1.14).

In this case, the oracle seems not to have supplied information relevant to assessing the likelihood of a given outcome. Rather, the two disputing parties agreed to designate the oracle as an arbitrator. It is accepted by both sides because of its assumed impartiality, and empowered to make a choice between two alternatives that will be binding on the contestants. Arbitration and the importance of impartiality to the arbitrator's success in solving bargaining problems will be a recurring theme in later chapters.

Although Herodotus provides no details, coming to an agreement to accept arbitration is necessarily a difficult process (see further, chapters 3 and 4), raising questions of how it was that the many members of the pro-Kandaules faction agreed among themselves to abide by the oracle's decision and why each party regarded as credible the commitment of the other side to abide by the outcome. What is tolerably clear is that the resort to an arbitrator, because it was successful, avoided what might have been a debilitating conflict, ultimately destructive to both sides. And it is significant that the oracle was handsomely rewarded for its service: Herodotus' concern with the "arbitrator's payoff" in the process of state formation will figure prominently in chapter 3.

Plato alludes to the original emergence of structured social order, in the passage from book 2 of the *Republic* with which we began. There, Glaucon claims that social constraints on the pristine freedom of each individual to act so as to gain "more and more" were established by an agreement (*sunthêkai*) among mutually fearful persons. The question we are now faced with, as in the story of the Delphic solution to the Lydian civil war, is how people can come to that sort of agreement, given the conditions of rationality that emerge from the story of Gyges. That is the subject of the next chapter.

But before leaving the paired tales of Gyges the Lydian, there is a final puzzle that we might hope to solve: Why, assuming that Plato adapted his version of the story from that of Herodotus, did he move the story of seduction and usurpation back one or more generations, from "Gyges" to "an ancestor of Gyges"? Herodotus' tale of Gyges concludes (1.13.2) with the comment that, after the Delphic oracle had awarded the throne to Gyges, it warned that the relatives of Kandaules would have vengeance on Gyges' descendants "in the fifth generation." Herodotus notes that the Lydians and their kings paid no attention to this (that is, they discounted future costs relative to current gains). But, he continues, the oracle was fulfilled with the

defeat by Darius of Persia of Gyges' great-great grandson, Croesus. Herodotus' story of Gyges is framed, beginning and end, by the shadow of divine retribution. As we have seen, in Plato's version of the story, Adeimantus pointedly insists that the opportunity for preference satisfaction of the unconstrained rational man, modeled by the holder of a ring of invisibility, must not be compromised by any concern for divine retribution. By pushing the responsibility for the act of regicide that founded a new dynasty back to "an ancestor of Gyges," who could not, therefore, have known the "fifth-generation" oracle delivered to Herodotus' Gyges, the donor of Delphi's "Gygian treasure," Plato confounded the "fifth-generation" oracle. Thus he dispelled whatever shadow a divinely ordained future might have cast on the preferences or the payoffs to the anonymous, fully informed, fully rational, egoistically self-interested subject of his carefully plotted tale.

In his story of Gyges, Herodotus offered an account of seemingly rational actors with ranked (if not fully revealed) preferences, and coherent (if not always accurate) beliefs. Herodotus' Gyges made his highly constrained life-or-death choices based on the choices he had reason to believe would be made by other agents, each seeking his and her own goals.⁷⁴ Herodotus' readers thus confront a situation readily rendered as a strategic game. Plato, writing a generation after Herodotus, adapted the now-familiar familiar Gyges story to his own purposes; transforming it, by the device of the magic ring, into a readily-solved parametric decision problem. He was able to do so because, even with its allusion to the necessities of fate, Herodotus' story about Gyges was already framed around the choices of instrumentally rational actors: both Herodotus and Plato assumed readers who were familiar with what I have been calling the background Greek folk theory of practical reasoning. As we will see in the following chapters, Herodotus and Plato were not exceptional within their intellectual milieu in presenting situations of choice in forms that are readily reduced to decision-trees and strategic games. Nor were they exceptional in offering their readers models of human rationality that remain applicable to social behavior in the real world.

Figure 1.1. Plato, *Republic*, book 2. Gyges and the ring decision tree.
 N = Nature. G = Gyges.

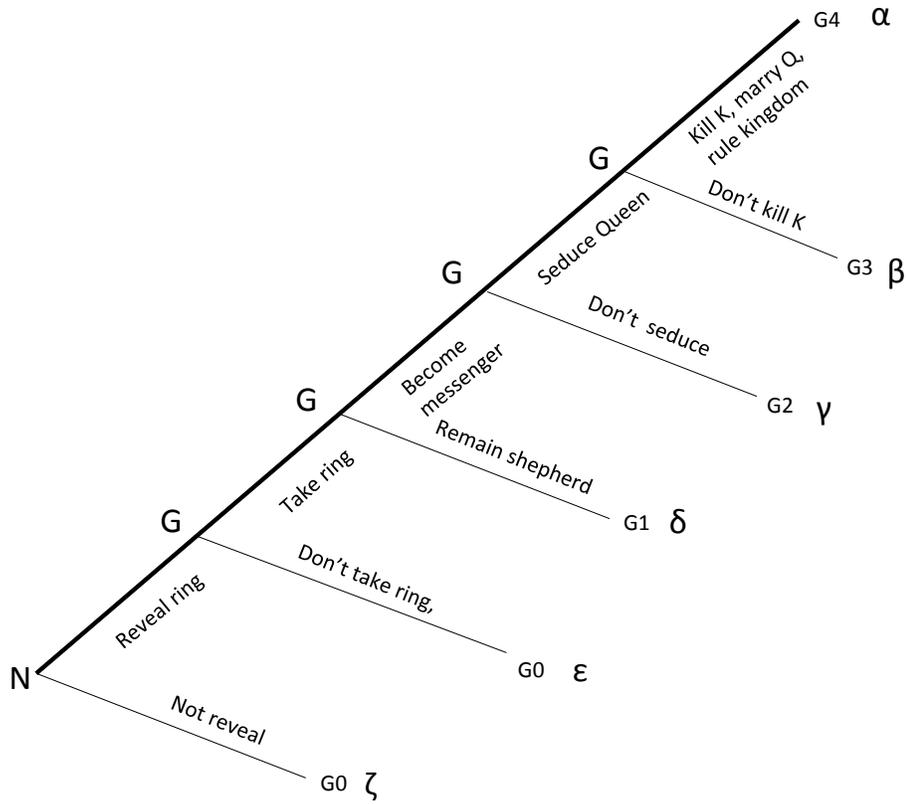


Figure 1.2. Aristotle, *Prior Analytics*. Example of preference ranking.

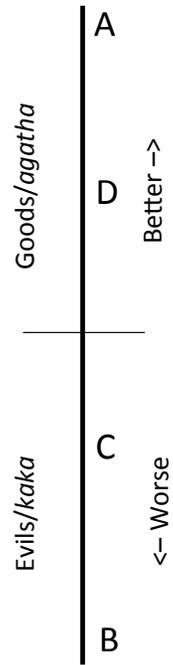


Figure 1.3. Herodotus, *Histories*, book 1, Gyges and Queen Subgame.
G = Gyges. Q = Queen.

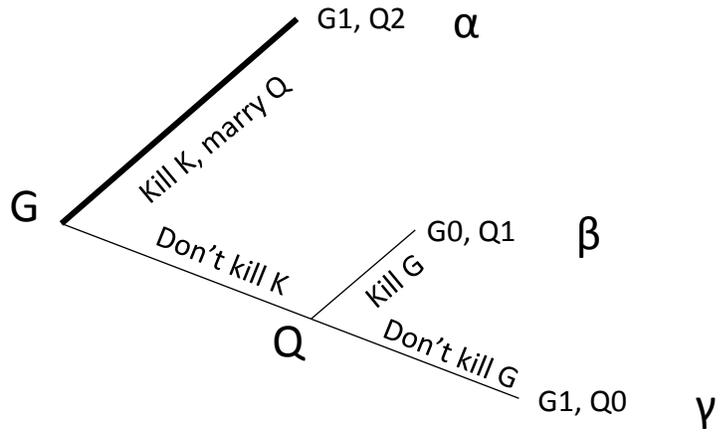
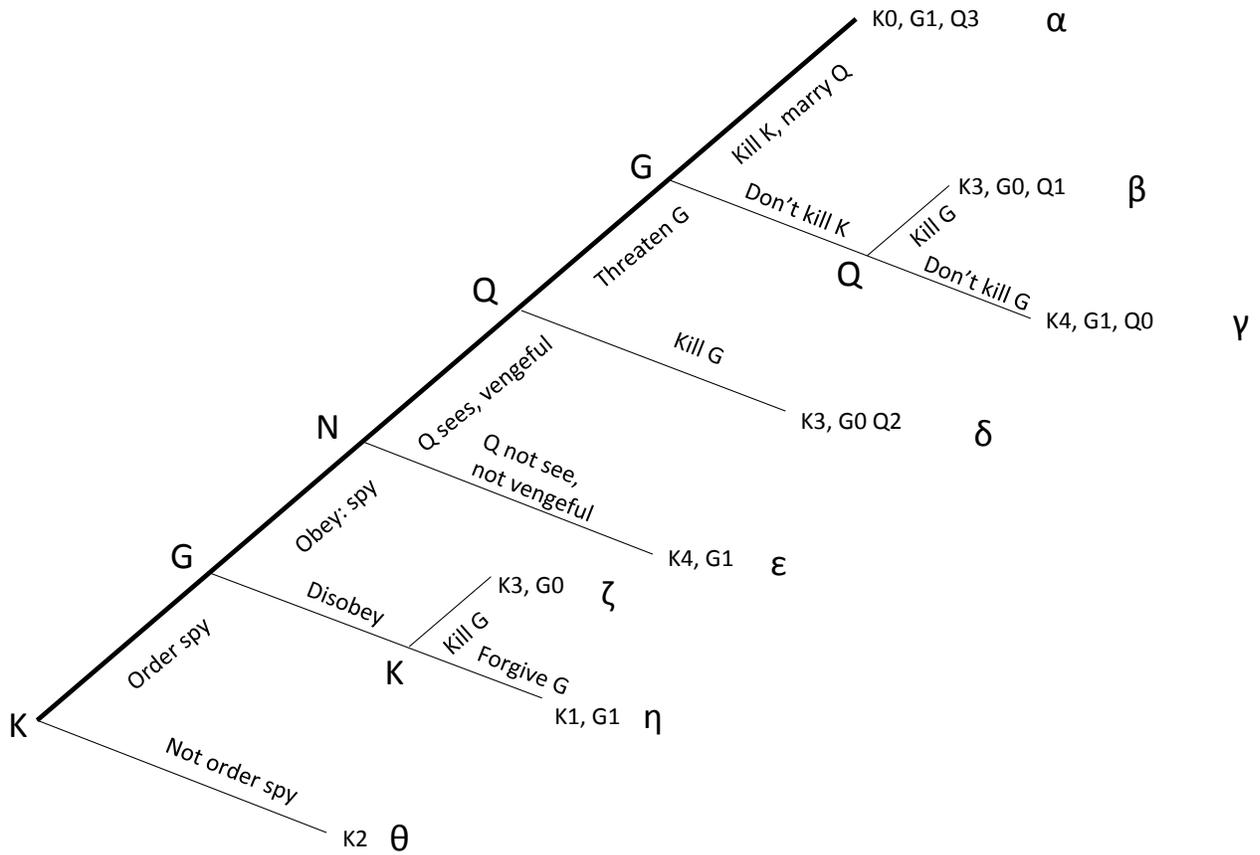


Figure 1.4. Herotodus, *Histories*, book 1, King, Gyges, and Queen Game.
 K = King, G = Gyges, N = Nature, Q = Queen.



1. Gyges. References

- Adam, James and D.A. Rees. 1963. *Plato, The Republic [with critical notes and commentary]*. Cambridge: Cambridge University Press.
- Allen, James. 1993. "Failure and expertise in the ancient conception of an art." Pp. 83-110 in *Scientific Failure*, edited by Allen I. Janis.: Rowman & Littlefield Publishers.
- Amadae, S. M. 2003. *Rationalizing capitalist democracy: The Cold War origins of rational choice liberalism*. Chicago: University of Chicago Press.
- Amadae, S. M. 2016. *Prisoners of reason: Game theory and neoliberal political economy*.
- Anderson, Greg. 2018. *The Realness of Things Past: Ancient Greece and Ontological History*. New York: Oxford University Press.
- Annas, Julia. 1981. *An introduction to Plato's Republic*. Oxford: Clarendon Press.
- Asheri, David, Alan B. Lloyd, Aldo Corcella, Oswyn Murray, Alfonso Moreno, and Maria Brosius. 2007. *A commentary on Herodotus: Books I-IV*. Oxford and New York: Oxford University Press.
- Balot, Ryan K. 2001. *Greed and injustice in classical Athens*. Princeton, N.J.: Princeton University Press.
- Bermudez, Jose. 2015. "Strategic vs. Parametric choice in Newcomb's Problem and the Prisoner's Dilemma: Reply to Walker." *Philosophia*. 43:xxx-xxx.
- Bernstein, Peter L. 1996. *Against the gods: The remarkable story of risk*. New York: John Wiley & Sons.
- Binmore, K. G. 2007. *Game theory: A very short introduction*. Oxford and New York: Oxford University Press.
- Buchak, Lara Marie. 2013. *Risk and rationality*. Oxford: Oxford University Press.
- Chwe, Michael Suk-Young. 2013. *Jane Austen, Game Theorist*: Princeton University Press.
- Danzig, Gabriel. 2008. "Rhetoric and the Ring: Herodotus and Plato on the Story of Gyges as a Politically Expendient Tale." *Greece & Rome*. 55 (2):169-192.
- Davis, Morton D. 1983. *Game theory: A nontechnical introduction*. New York: Basic Books.
- Dixit, Avinash K. and Susan Skeath. 1999. *Games of Strategy*. New York: W. W. Norton & Company.
- Dodds, E. R. 1959. *The Greeks and the irrational*. Berkeley, Calif.: University of California Press.
- Ferejohn, John. 2009. "Expectations, Institutions and Rationality in Political Theory." in *Gurobaru-shakai*, edited by Masaru Kohno.: Keiso Shobo.
- Hardin, Russell. 2007. *David Hume: Moral and political theorist*. Oxford and New York: Oxford University Press.
- Irwin, Terence H. 1999. "Republic II: Questions about Justice." Pp. 164-185 in *Plato 2. Ethics, Politics, Religion, and the Soul*, edited by Gil Fine. Oxford: Oxford University Press.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Katemkar, Rachel. n.d. "Plato's Statesman on Measurement in Politics." Paper draft

- presented at Plato and Rationality Conference, Stanford, March 2019.
- Kirwan, Christopher. 1965. "Glaucón's Challenge." *Phronesis*. 10 (2):162-173.
- Laird, Andrew. 2001. "Ringing the Changes on Gyges: Philosophy and the Formation of Fiction in Plato's *Republic*." *Journal of Hellenic Studies*. 121:12-29.
- Laks, André and Glenn W. Most, ed. 2016. *Early Greek philosophy VIII: The Sophists*. Vol. 1. Cambridge, Massachusetts: Harvard University Press.
- Myerson, Roger B. 2009. "Learning from Schelling's *Strategy of Conflict*." *Journal of Economic Literature*. 47 (4):1109-1125.
- Schelling, Thomas C. 1980. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Sen, Amartya. 2017. *Collective choice and social welfare*. Cambridge, Mass.: Harvard University Press.
- Shields, C. 2006. "Plato's Challenge: The Case Against Justice in *Republic II*." Pp. 63-83 in *The Blackwell Guide to Plato's Republic*. Malden, MA: Blackwell.
- Simon, Herbert Alexander. 1955. "A behavioral model of rational choice." *Quarterly Journal of Economics*. 65:99-118.
- Stone, Peter. 2004. "Review of Amadae 2003." *Perspectives on Politics*. 2 (2):347-348.
- Taylor, C.C.W. 2007. "Nomos and Physis in Democritus and Plato." *Social Philosophy & Policy*. 24:1-20.
- Weiss, Roslyn. 2007. "Wise Guys and Smart Alecks in *Republic 1* and *2*." Pp. 90-115 in *The Cambridge Companion to Plato's Republic*, edited by G.R.F. Ferrari.

1. Gyges. Notes (still vestigial and unedited).

¹ Par ma foi, il y a plus de quarante ans que je dis de la prose sans que j'en susse rien.

² Sun Tzu, *Art of War*; Kautilya, *Arthashastra*. Examples could certainly be multiplied.

³ Cf. Schelling 1960: 16-17: "Furthermore, theory that is based on the assumption that the participants coolly and 'rationally' calculate their advantages according to a consistent value system forces us to think more thoroughly about the meaning of 'irrationality' ... departures from complete rationality may be in many different directions...It may not be an exaggeration to say that our sophistication sometimes suppresses sound intuitions, and one of the effects of an explicit theory [of rationality] may be to restore some intuitive notions that were only superficially 'irrational.'"

⁴ For a strong and explicit statement of this approach see G. Anderson 2018.

⁵ Cf. Amadae 2003 (with Stone 2004), 2016 and more generally critique of instrumental reason by Adorno, Horkheimer, and others of the Frankfurt School xx.

⁶ E.g. *hairesis*, *anangke*, *to sumpheron*, *lusiteleia*, *eikos* respectively.

⁷ Cf. Myerson 2009 on the way in which analytic, but informal work on conflict by Thomas Schelling (1960) both anticipated and stimulated the development of formal game theory.

⁸ Trans. Laks and Most, Loeb, adapted. Full passage: προσερωτώμενος δέ, εἰ τοὺς ἐπισταμένους μὲν ἃ δεῖ πράττειν, ποιῶντας δὲ τάναντία σοφούς τε καὶ ἐγκρατεῖς εἶναι νομίζοι, οὐδέν γε μᾶλλον, ἔφη, ἢ ἀσόφους τε καὶ ἀκρατεῖς· πάντα γὰρ οἶμαι προαιρουμένους ἐκ τῶν ἐνδεχομένων ἃ οἴονται συμφορώτατα αὐτοῖς εἶναι, ταῦτα πράττειν. NB the likely textual problem with ἀκρατεῖς in the first sentence: in the second phrase, for ἀκρατεῖς read instead ἐγκρατεῖς. Thus emended, the contrast is between "clever and willful" (the "Thrasymachean" take on egoistic choice-making, on which, see below) and "unwise and willful" (the Socratic paradox that no one does wrong willingly) – the point being that *akrasia* is an empty concept for Socrates. Xenophon's Socrates is, in some ways, quite different from Plato's character of that name (for example he seems much more focused on practical matters of household and polis management), but he shares some of the same concerns about human motivation.

⁹ Depending on how we read the force of προ- in the participle προαιρουμένους, it may be that Socrates means that if and when people engage in a process of deliberating in advance on how they ought to choose, and when they base their

deliberations on an assessment of what is available to them, then (and perhaps only then), they make their choice based on what they think is most advantageous to themselves, and act accordingly. My thanks to Terence Irwin for discussion of this point.

- ¹⁰ On this account, what is surprising about Xenophon's Socrates' statement is not its content as such, but rather its deployment in a discussion of the question of *akrasia*.
- ¹¹ Practical reasoning, in the sense employed here, concerns motivations and is aimed at explaining actions: why and how people go for what they do go for; it is related to, but distinct from, moral reasoning about the value of ends and ethical reasoning about the rightness of the means employed. By "folk theory," I mean that it is a theory that is common to an intellectual community but has no definable author. This is not "The Folk Theorem" of repeated Prisoner's Dilemma games, discussed by Binmore 2007:75-76, but my use of the term "folk theory" is inspired by that rubric.
- ¹² Normative and descriptive: In introducing their "prospect theory" critique (see ch. 6), Kahneman and Tversky 1979: 263, write: "Expected utility theory has dominated the analysis of decision making under risk. It has been generally accepted as a normative model of rational choice, and widely applied as a descriptive model of economic behavior, e.g. Thus, it is assumed that all reasonable people would wish to obey the axioms of the theory, and that most people actually do, most of the time." Hardin 2007 offers a succinct introduction.
- ¹³ Ordered preferences are complete, ranked, and transitive, meaning that if there are three possible outcomes (A,B,C), then A is preferred to B, B to C, and A to C.
- ¹⁴ While reformulating the folk theory in terms of his own theory of knowledge, Plato's Socrates employs the ordinary assumptions of the folk theory as a sort of intuition pump for his interlocutors: See, for example Plato *Protagoras* 358c-d: Socrates claims that no one goes willingly (*hekon*) toward evils (*kaka*), or what he considers evil, neither is it within human nature (*anthropou phusis*) to wish to go towards what one thinks to be evil rather than good. "And when one is forced to choose between two evils, no one will choose the greater [evil] if the lesser [evil] is available." For Plato's Socrates' association of "goods and evils" with "pleasures and pains," see Plato, *Protagoras* 357d-e: People make mistakes with regard to the choice of pleasures and pains (*ton hedonon ... kai lupon*) – that is (*tauta de esti*), with regard to *agatha* and *kaka* – and make these mistakes out of lack on knowledge. I use the masculine form here and elsewhere when referring generically to ancient thought, because the standard individual assumed by most ancient Greek social theorists was male. I use the feminine form elsewhere.
- ¹⁵ The instrumentally rational agent might choose an option that is regarded by the agent as a good in itself. That is, instrumental rationality does not preclude the

belief that some things – e.g. justice, virtue – are intrinsically valuable. But it is the expected advantage to the agent, not the inherent goodness of the option, that is the reason for the agent’s choice. See further, chapter 8.

- ¹⁶ On the standard theory of expected utility maximization, and its limitations, see Buchak 2013.
- ¹⁷ I have tended to use the terms “advantage (to the agent),” or “(subjective) value,” or “(subjective) happiness” to refer to the “satisfaction with outcomes” that is maximized by the choice-making agent, but I do not mean to imply that practical reasoning must be grounded in an ethics of hedonism (as pleasure-seeking); see chapter 7 on Greek eudaimonism. For a very clear sketch of preferences, beliefs, expectations, and actions in contemporary rational choice theory, see Ferejohn 2009. Measurement in Plato’s political theory: Katemkar in progress.
- ¹⁸ NB here I refer to Thrasymachus of the Republic. Contrast the fragments.
- ¹⁹ A striking example of that line of argument is preserved in the fifth-century (or possibly fourth-century) text known to as the “Anonymous Iamblich.” xx
- ²⁰ τὸν ἄρχοντα, καθ’ ὅσον ἀρχῶν ἐστίν, μὴ ἀμαρτάνειν, μὴ ἀμαρτάνοντα δὲ τὸ αὐτῷ βέλτιστον τίθεσθαι. On the conception of a complete “art” that must account for the possibility of failure that is implied in this passage see Allen 1993: 87-89: the true craftsman may fail in his results, but he never fails as a craftsman, qua craftsman, or at his craft. As Allen points out, this places the full onus on the performance of the craftsman; other ancient conceptions of failure appeal to external circumstances or luck.
- ²¹ οὐδέποτ’ ἄρα... λυσιτελέστερον ἀδικία δικαιοσύνης:
- ²² Θρασύμαχος γάρ μοι φαίνεται πρωαίτερον τοῦ δέοντος ὑπὸ σοῦ ὥσπερ ὄφεις κηληθῆναι. 2.358b.
- ²³ καὶ πρῶτον μὲν ἐρῶ δικαιοσύνην οἷον εἶναί φασιν καὶ ὄθεν γεγονέναι, δεύτερον δὲ ὅτι πάντες αὐτὸ οἱ ἐπιτηδεύοντες ἄκοντες ἐπιτηδεύουσιν ὡς ἀναγκαῖον ἀλλ’ οὐχ ὡς ἀγαθόν, τρίτον δὲ ὅτι εἰκότως αὐτὸ δρῶσι: πολὺ γὰρ ἀμείνων ἄρα ὁ τοῦ ἀδίκου ἢ ὁ τοῦ δικαίου βίος, ὡς λέγουσιν. ἐπεὶ ἔμοιγε, ὦ Σώκρατες, οὐ τι δοκεῖ οὕτως: ἀπορῶ μέντοι διατεθρυλημένος τὰ ᾧτα ἀκούων Θρασυμάχου καὶ μυρίων ἄλλων.
- ²⁴ On the question of why the “new beginning” of book II is called for, and what Glaucon and Adeimantus ask of Socrates, see Kirwan 1965; Irwin 1999.
- ²⁵ On the challenge offered by Glaucon and Adeimantus and the philosophical purposes to which the story of Gyges is put by Plato, see Shields 2006; Weiss

2007; Taylor 2007: 13-19; Reeve 2008. Annas 1981: 64-71, argues that “We live in a world where we have to take account of the natural and artificial consequences of injustice, and it is merely silly to ask what we would do if we escaped these by having magic rings” (p. 69). She concludes that by the inclusion of the Gyges story, Plato shows that his theory of justice is not meant to be realistic. While the *Republic* is certainly not meant to be merely realistic, I think Annas misses the point of the Gyges story as a thought experiment aimed at separating what she describes as “natural and artificial consequences.”

²⁶ ὡς δὲ καὶ οἱ ἐπιτηδεύοντες ἀδυναμία τοῦ ἀδικεῖν ἄκοντες αὐτὸ ἐπιτηδεύουσι, μάλιστα ἂν αἰσθοίμεθα, εἰ τοιόνδε ποιήσαιμεν [359ξ] τῆς διανοίας: δόντες ἐξουσίαν ἑκατέρῳ ποιεῖν ὅτι ἂν βούληται, τῷ τε δικαίῳ καὶ τῷ ἀδίκῳ, εἴτ' ἐπακολουθήσαιμεν θεώμενοι ποῖ ἢ ἐπιθυμία ἑκάτερον ἄξει. ἐπ' αὐτοφώρῳ οὖν λάβοιμεν ἂν τὸν δίκαιον τῷ ἀδίκῳ εἰς ταύτῳ ἰόντα διὰ τὴν πλεονεξίαν, ὃ πᾶσα φύσις διώκειν πέφυκεν ὡς ἀγαθόν.

²⁷ Compare Davis 1983: 61: “The game theorist, to paraphrase Lewis Carroll, must pick the proper road after being told the player’s destination; and players don’t need to know anything about game theory, but they have to know what they like.” On the road as a metaphor for the kind of life one leads in Plato and elsewhere in Greek literature, see Laird 2001: 16-17.

²⁸ On the meaning of *pleonexia* in classical thought see Balot 2001.

²⁹ Literature on Plato’s Gyges story: Adam and Rees (1962); Laird 2001; Danzig 2008.

³⁰ Many commentators have taken the protagonist of the book 2 narrative to be named Gyges on the basis of *Republic* 10.612b: Socrates refers to “the ring of Gyges” (*ton Gugou daktulion*). See however, Adam and Rees (1962: 126-27 = Book 2 appendix 1), for the difficulty and various possible solutions. See further Laird 2001: xx, Danzig 2008: xx. I will suggest, below (section 1.9), a reason that Plato might have chosen to speak of an “ancestor of Gyges” here in book 2.

³¹ Note that Plato first offers an accessible narrative, followed by abstraction: the compelling narrative about a person with a background story (Gyges the Lydian shepherd), immediately followed by a discussion of abstract agents (the “two men”). The reader is thereby dissuaded from reading anything into Gyges’ background by, for example, imagining that his motivation is specific to someone in a certain social role. My thanks for Demetra Kasimis for pushing me on this point.

³² καὶ τᾶλλα πράττειν ἐν τοῖς ἀνθρώποις ἰσόθεον ὄντα. Note that being *isotheos* “equal to a god” is comparable to being *eudaimon*, literally “well-godded.” Here the reference is clearly to subjective preference satisfaction.

- ³³ οὕτω δὲ δρῶν οὐδὲν ἄν διάφορον τοῦ ἐτέρου ποιοῖ, ἀλλ' ἐπὶ ταῦτ' ἄν ἴοιεν ἀμφοτέρω.
- ³⁴ Why the Queen, having been somehow seduced by an invisible man, then helped him to kill her husband, and how Gyges then gained the kingship remain unexplained in Plato's story. Contrast Herodotus' story of Gyges, below, where the Queen's motivation is made clear.
- ³⁵ Of course this begs the question of how Glaucon knows the story of Gyges' coming to power. But this is not a problem we need to worry about if we take Glaucon at his word: he is not writing history, but developing a thought experiment. The unnamed ancestor of Gyges is a stand-in for the perfectly unjust man, not a historical figure that we ought to praise or blame. And see below, on Herodotus' story. Gyges and Plato's probable adaptation of it. Note also that the stipulation that the unjust man will be beloved of the gods again tracks a subjective-happiness construal of *eudaimonia*.
- ³⁶ $X > Y > Z$ (no cycles, so $Z < X$). Note that $>$ is, in practice, \geq .
- ³⁷ ἀθλιώτατος μὲν ἄν δόξειεν εἶναι τοῖς αἰσθανομένοις καὶ ἀνοητότατος.
- ³⁸ πρῶτον μὲν οὖν ὁ ἀδίκος ὥσπερ οἱ δεινοὶ δημιουργοὶ ποιεῖτω ... τὰ τε ἀδύνατα ἐν τῇ τέχνῃ καὶ τὰ δυνατὰ διαισθάνεται.
- ³⁹ E.g. Dixit and Skeath 1999: 27: "rationality has two essential ingredients: complete knowledge of one's own interests, and flawless calculation of what action will best serve those interests."
- ⁴⁰ Per note 24, above, the motivations of the Queen in becoming his collaborator in killing the King, are unspecified in Plato's story, and do not affect Gyges freedom. Her role is very different in Herodotus' story of Gyges; see below.
- ⁴¹ Revealed preferences: Binmore 2007: 6-7.
- ⁴² My thanks to Avidit Acharya for help with this hypothetical.
- ⁴³ This problem of the risk associated with the likelihood (or probability) of an outcome will be treated in more detail below. Suffice it to say here that what seems excessively risky to one choice-maker might not seem so to another.
- ⁴⁴ ἐπ' αὐτοφώρῳ οὖν λάβοιμεν ἄν τὸν δίκαιον τῷ ἀδίκῳ εἰς ταύτῳ ἰόντα διὰ τὴν πλεονεξίαν, ὃ πᾶσα φύσις διώκειν πέφυκεν ὡς ἀγαθόν,
- ⁴⁵ See now Melissa Lane, Carlyle Lectures book on Greek offices, forthcoming.

⁴⁶ Game theory on diversity of preferences: Davis (1983); 61: “In fact, we cannot make any general assumption about people’s wants, because different people want different things. Ibid: 62: “A utility function is simply a “quantification” of a person’s preferences with respect to certain objects.”

⁴⁷ On the difference between parametric choice situations and strategic games see Bermudez 2015: 7 (ms): “In parametric choice situations agents make their choices with the background fixed. The only dimension of variation is the agent’s own choice. The classically rational decision-maker therefore seeks to maximize expected utility relative to parameters that are set by the environment, and her choice depends only on her preferences and probability assignments. ... As Jon Elster puts it, “In the strategic or game-theoretic mode of interaction, each actor has to take account of the expectations of all other actors, including the fact that their actions are based upon their expectations concerning his own” (Elster 1979, p.18). But when the outcome can be identified by specifying one’s own choice, no such taking account of the expectations of other participants is required. The other participants are simply pieces of the environment, and so these situations are better viewed as instances of parametric choice.”

⁴⁸ Όταν δὲ δυοῖν ὄντων τὸ Α τοῦ Β αἰρετώτερον ἢ, ὄντων ἀντικειμένων, καὶ τὸ Δ τοῦ Γ ὡσαύτως, εἰ αἰρετώτερα τὰ ΑΓ τῶν ΒΔ, τὸ Α τοῦ Δ αἰρετώτερον.

⁴⁹ We could (as Aristotle did not) assign quantities to each condition according to Aristotle’s specifications: A=4, D=3, C=2, B=1. Per the formula above: $4 > 1$, $3 > 2$, $(4+2) > (1+3) \rightarrow 4 > 3$, $2 > 1$. Note, however that there are other cases in which the conditions are satisfied but A is not preferable to (greater than) D. For example, A=5, B=1, C=4, and D=6. So in this latter case, *sunousia* (D) would be preferred to *philia* (A), falsifying Aristotle’s conclusion. So either Aristotle did not think of this problem in terms of quantities (thus *haireteron* is not to be thought of as “greater than in a generally quantifiable sense”), or he has offered a false result. My thanks to Peter Stone for pointing this out.

⁵⁰ Εἰ δὴ ἔλοιτο πᾶς ὁ ἐρῶν κατὰ τὸν ἔρωτα τὸ Α τὸ οὕτως ἔχειν ὥστε χαρίζεσθαι καὶ τὸ μὴ χαρίζεσθαι τὸ ἐφ’ οὗ Γ, ἢ τὸ χαρίζεσθαι τὸ ἐφ’ οὗ Δ καὶ τὸ μὴ τοιοῦτον εἶναι οἷον χαρίζεσθαι τὸ ἐφ’ οὗ Β, δῆλον ὅτι τὸ Α τὸ τοιοῦτον εἶναι αἰρετώτερόν ἐστιν ἢ τὸ χαρίσασθαι. τὸ ἄρα φιλεῖσθαι τῆς συνουσίας αἰρετώτερον κατὰ τὸν ἔρωτα. μᾶλλον ἄρα ὁ ἔρως ἐστὶ τῆς φιλίας ἢ τοῦ συνεῖναι· εἰ δὲ μάλιστα τούτου, καὶ τέλος τοῦτο. τὸ ἄρα συνεῖναι ἢ οὐκ ἔστιν ὅλως ἢ τοῦ φιλεῖσθαι ἔνεκεν· καὶ γὰρ αἱ ἄλλαι ἐπιθυμίαι καὶ τέχναι οὕτως.

⁵¹ Note that there seems to be a missing term E: that is having the love of the beloved and having sex with the willing beloved. It seems on the face of it that Aristotle thinks E would be better than A (since $D > C$) but E is not interesting in re. getting the ranking of *philia* and *sunousia* right, so E is left out of the ranking.

- ⁵² Behavioral economics followed soon upon the original program of rational choice and game theory: Simon 1955. Behavioral psychology: Kahneman 2011.
- ⁵³ Cf. Laird 2001: 18: “the kind of crises of moral choice which involve Candaules, his wife and Gyges in Herodotus are, in Glaucon's story more instructively centred on one person.” This accurately captures the difference between the 1-player and multi-player version of the story, but puts too much emphasis on morality. Danzig 2008: 174 correctly notes that “The term *hairesis* (choice), which appears [in Herodotus' story] twice as a noun and once as a verb (1.11.2,3,4), does not signify a specifically moral choice for which Gyges is to be called to account but a practical choice between two alternatives.”
- ⁵⁴ Herodotus' Gyges story, with literature review: Asheri et al (2007): 81-83.
- ⁵⁵ ἀρρωδέων μὴ τί οἱ ἐξ αὐτῶν γένηται κακόν.
- ⁵⁶ ὁ μὲν δὴ ὡς οὐκ ἐδύνατο διαφυγεῖν, ἦν ἔτοιμος. 1.10.1.
- ⁵⁷ ἀρχὴν γὰρ ἐγὼ μηχανήσομαι οὕτω ὥστε μηδέ μαθεῖν μιν ὀφθεῖσαν ὑπὸ σεῦ. 1.9.1
- ⁵⁸ σοὶ μελέτω τὸ ἐνθεῦτεν ὅκως μὴ σε ὄψεται ἰόντα διὰ θυρέων. 1.9.3
- ⁵⁹ νῦν τοί δυῶν ὀδῶν παρεουσέων Γύγη δίδωμί ἀίρεσιν.
- ⁶⁰ ἰκέτευε μὴ μιν ἀναγκαίη ἐνδέειν διακρῖναι τοιαύτην ἀίρεσιν.
- ⁶¹ ἀλλ' ὦρα ἀναγκαίην ἀληθέως προκειμένην ἢ τὸν δεσπότεα ἀπολλύναι ἢ αὐτὸν ὑπ' ἄλλων ἀπόλλυσθαι: αἰρέεται αὐτὸς περιεῖναι.
- ⁶² Other literature on the relationship between the two stories and their *Nachleben*: Smith 1920; Laird 2001: 29 with n. 59.
- ⁶³ Cf. note 17, above.
- ⁶⁴ King is indifferent between γε: 4, βδζ: 3. Gyges is indifferent between αση: 1, βδζ: 0.
- ⁶⁵ If, for example, a player is indifferent between getting outcome A and a 1:10 chance of getting outcome B, then he prefers B to A by a factor of 10. If we have set the payoff for A at 1, then we know his payoff for B is 10.
- ⁶⁶ On cluelessness in game theory, see Chwe 2013.
- ⁶⁷ Gyges' unusual passion for his wife: Laird 2001.

⁶⁸ μήτε γυναῖκα τὴν ἐμήν, μὴ τί τοι ἐξ αὐτῆς γένηται βλάβος.

⁶⁹ If, counterfactually, Herodotus had assigned a percentage to Kandaules' assessment of the probability of the combination "Gyges discovered, the Queen seeks vengeance on the King," we would be able to assign cardinal weights to Kandaules' preferences. Suppose that Herodotus had written that Kandaules believed that there was a one in ten chance of Gyges being spotted. This would mean that (momentarily ignoring the variables of the Queen's vengefulness and the object of her revenge) we could calculate the relative weights of α and θ . Arbitrarily setting θ , as we have, at 2, and assuming that there is a point at which K is indifferent between the outcomes of dying having spied and living without spying, we could calculate the value of α for K at 0.2. Since the difference between 2 and 0.2 seems likely to underestimate K's value on his own life, we might suppose that K regards the likelihood of α as 1:1000 so α should be set at 0.002. Now we could bring back in the vengefulness of the Queen and the object of her vengeance, and suppose that a rational K estimated the chance of G being seen by Q as one in a 10, the chance of the Queen being murderously vengeful at 1 in 10, and the chance that the King would be the object of her vengeance at 1 in 10; thus the chance that his life was at risk is 1:1000. Assuming that he valued his life highly (that is, α is closer to 0.002 than 0.2) and was not just terribly unlucky (always possible in a lottery), then his mistake (if he made one) lay in underestimating the likelihood of the full sequence, "Gyges spotted, the Queen vengeful, and himself the object of her ire." That is the same conclusion that we arrived at above, without numerical estimates of probability or cardinal preference weighting.

⁷⁰ ἐρασθεὶς δὲ ἐνόμιζέ οἱ εἶναι γυναῖκα πολλὸν πασέων καλλίστην 1.8.1; γὰρ ἐγὼ μηχανήσομαι. 1.9.1

⁷¹ χρῆν γὰρ Κανδαύλη γενέσθαι κακῶς.

⁷² The stark contrast between ancient practices of oracular divination and modern assessments of risk is, for example, the theme of Bernstein 1996.

⁷³ ἔσχε δὲ τὴν βασιληίην καὶ ἐκρατύνη ἐκ τοῦ ἐν Δελφοῖσι χρηστηρίου.

⁷⁴ Without any reference to rational choice or game theory, Dewald 2003: 41-42, 49 makes this point explicitly: "What the self-contained Herodotean *logos* focuses on is the immediate context, an individual's goals, and his or consequent actions within that *logos*. Whether the individual depicted is a poor herdsman, someone's wife or mother, royal courtier, Greek citizen, prostitute, eastern potentate, or Sicilian adventurer, his or her actions in the Herodotean narrative are first defined by the constraints of external circumstances. Within those constraints, he or she is shown acting out of a desire to obtain some particular objective through

whatever means are realistically available... Each man or woman in Herodotus wants as much scope for following his or her personal objectives as he or she can get. Each one, however, is ultimately held in check by the existence of all the others and their competing and contradictory desires.”