

MAKING GOOD CHOICES: AN INTRODUCTION TO PRACTICAL REASONING

CHAPTER 14. IRRATIONAL CHOICES: SOME COMMON FALLACIES OF PRACTICAL REASONING

There are many ways to reason poorly, unfortunately, and there are many factors that can cause this. In the field of critical reasoning, a number of interesting patterns of bad reasoning have long been identified. These patterns are called common (logical) fallacies, “common” because they are so frequently committed and also because they are often mistakenly accepted as perfectly good ways to reason. Here, “fallacy” does not mean false. In this special use, “fallacy” means a mistake or error of reasoning. A fallacy is a failure of reasoning to conform to the appropriate norm or standard of rationality. Because a fallacy is a failure of rationality, it is an instance of irrationality. Along with being interesting in their own right, these common critical reasoning fallacies are interesting because they reveal things about human nature. There are aspects of human nature, it seems, that create a tendency for us to make mistakes when we reason critically, and these parts of our nature that move us toward irrationality just when we are trying to be rational are important to know about. These common fallacies are studied so that people who are trying to reason critically are better able to avoid them.

In the area of practical reasoning there are, likewise, common errors that we humans seem prone to make. In this chapter we will identify these patterns of poor decision making: common **practical reasoning fallacies**. They all have a common general form: an agent has a goal and the agent’s menu of available options contain (at least) one that, if chosen, best achieves the goal compared to the other options, and the agent fails to choose it. As with the common logical fallacies of critical reasoning, these patterns of irrational choice seem to reveal interesting traits about human nature and how vulnerable we often are to reasoning errors. Aside from serving as a window into certain parts of human nature, these common patterns of irrational choice are widely studied so that, by becoming aware of them, agents are better able to avoid them.

3.1 Four categories of practical reasoning fallacies

It will help to locate practical reasoning fallacies if we first exclude cases of bad decisions that are not fallacies. Start with the broad category: decisions that fail to achieve the goal. We divide this large category into two groups: (a) there are decisions that fail due to events external to the agent that are beyond the agent's control. These are choices that have gone wrong, but are not necessarily bad choices. They typically involve failure of the state-of-the-world on which a decision's success depends. For example, a hired worker might forget to show up, or become ill, or even die. An object such as a computer, a car, or a piece of equipment might break or become unreliable. A system the agent "rents" such as the public electrical system, public transportation, or communication system might experience disruption or collapse. A corporation or a nation could have a decision ruined by an unusual weather condition or a natural disaster such as an earthquake. There are many examples of decisions that fail to result in goal achievement due to events external to the agent that are beyond the agent's control. The common fallacies of practical reasoning that we will be concerned with are not located in this group.

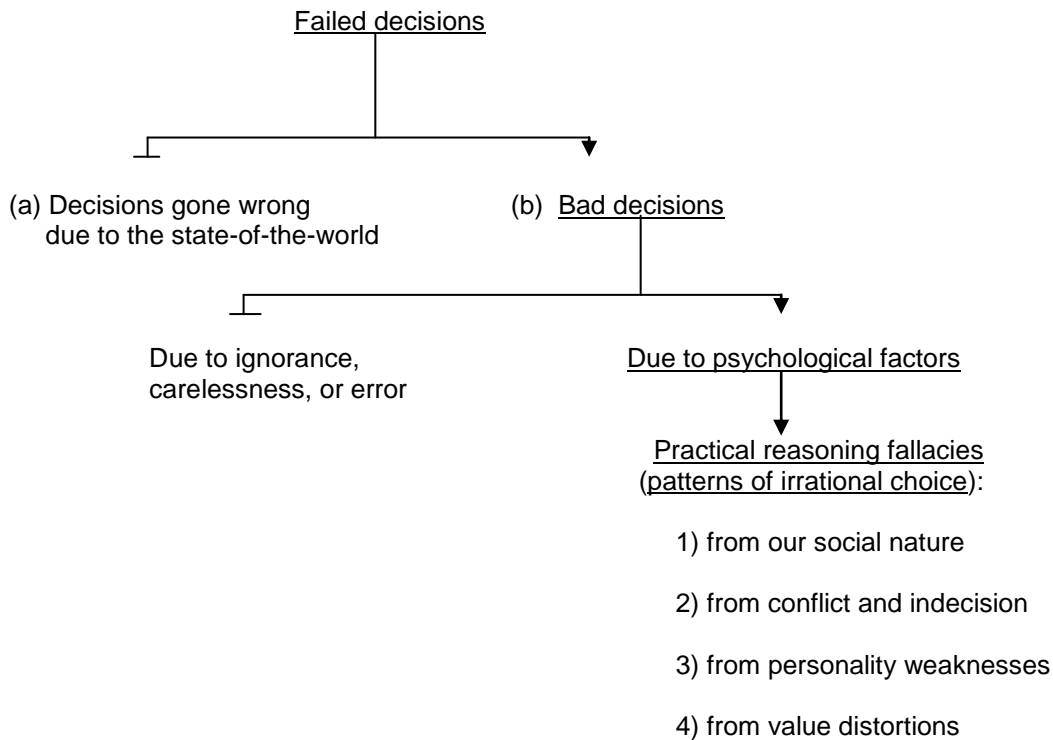
Aside from decisions that have gone wrong there are (b) decisions that fail due to events internal to the agent over which the agent has at least some control. These are, properly speaking, bad decisions. What can make a decision fail to result in goal achievement due to events internal to an agent, over which the agent can exercise at least some control? One possibility is that an agent might make a decision with insufficient knowledge. The knowledge that would have allowed the agent to make a better decision was available to the agent, let's suppose, but the agent did not make the effort to acquire the knowledge. Avoidable ignorance, then, is one possibility that can cause a bad choice.

Another possibility is that a bad decision results from carelessness on the agent's part. For example, an agent might form judgements too quickly and as a result they are false, or an agent might misperceive something because of poor observations that, had the agent been more careful in making observations, would have been perceived correctly. In these kinds of cases, decisions fail to achieve the goal because

of something internal to an agent – erroneous beliefs or misperceptions – but as with the case of insufficient knowledge the agent could have avoided such problems by her own efforts.

In each of these cases it is easy to see what the agent should do to avoid the cause of the bad decision, even if it is not always easy to do it. It is not always easy to acquire sufficient knowledge, or to form true beliefs, or to make careful observations. Yet it is clear in such cases that doing these things would remove the cause of such bad decisions. These, however, are not the patterns of irrational choice that are the common fallacies of practical reasoning. Fallacies are bad decisions of another kind, due to certain deeply rooted features of human nature.

There are often many psychological factors and events that are going on in an agent who is trying to come to a rational choice about what to do. For example, when making decisions any one of us might, and typically do, experience worries, fears, confusions, the force of habit, or the influence of personal bias or prejudice. Such psychological factors are not only happening *while* we deliberate about what to do, they have the power to *influence* decisions in a way that results in an irrational choice, a choice that the agent typically will regret having made. It is here that we located the common practical reasoning fallacies. These aspects of human nature are internal to the agent, and our assumption is that the agent is able to avoid, or at least minimize, their negative influence, even if it is not always easy to avoid the influence of things like fears, habits, or confusions. Let's put these features of human nature into four groups and for each describe the **practical reasoning fallacies** that commonly occur. There are (1) patterns of irrational choice that come from our social nature, (2) those that involve the experience of conflict and indecision, (3) those due to weak parts of our personality, and (4) those linked to distortion in what we desire and value. Before looking at each of these four kinds of bad decision making, here is a chart of the categories of failed decisions we have just covered.



3.2 Practical fallacies from our social nature

Poor decisions can result from the social nature of humans. Our social nature shows itself in how much we value our social relationships, in how much we desire to belong to certain social groups, and in how much of our lives are taken up with being members of various social institutions. Clearly, our “social identity” is a very large and important part of ourselves. We care about other people, some deeply and most others less deeply. And we want other people to care about us, some deeply and others less deeply. Because we care about what other people think and feel about us, we are motivated to do things that keep people accepting us; thus, we tend to do things that make people like us, and we refrain from doing things that would cause people to end their social connection to us. Typically, we don’t want our social relationships threatened nor do we want our social institutions to revoke our membership, and so we do things that cause other people to see us as socially acceptable and that give them reason to continue to accept us. Sometimes this works well, but at other times one or more well-known problems come up. The names of these problems are familiar and descriptive: problems of conformity, conventional wisdom, group-think, peer-pressure, and social authority. Why would a person desire to conform to social norms, customs and traditions? Why would a person feel pressure to follow peers? Why would

someone's thinking and belief system match that of the group or the broader conventional wisdom that everyone else accepts? Why are we prone to listen to and be influenced by the suggestions of those people whom we perceive to have high social status and authority? The general explanation is the same: humans are by nature social creatures; one's social connections to others is a fundamental part of being human.

How might our social nature interfere with making rational choices? This is the specific problem we want to look at, and we will focus on two ways this can happen: (a) by compromising an agent, and (b) by force of social habit.

(a) The **practical fallacy of decisional compromise**.

Suppose an agent has three options and wants to choose the one that maximizes goal achievement. Imagine that this agent's goal is reliable transportation and the agent has three cars from which to choose: C1, C2, and C3. Suppose that C2 is clearly the car that would provide the agent with the most reliable transportation and thus would be the rational choice, given the goal. And suppose that C1 is the least reliable transportation of the three, making it the most irrational choice of the three options. Now imagine that driving C2 makes the agent a non-conformist. Also, imagine that the agent's friends want him to drive C1 and exert "friendly" pressure for this choice. The agent also observes that many of those in positions of social authority, people of local influence to whom the agent looks up, drive C1 cars. Let's further say that it is conventional wisdom that driving C1 gives the image of success, and let's add that it is also considered patriotic, while C2 is not linked to success in peoples' minds and is even considered unpatriotic to drive. It is getting close to decision time for the agent. If the agent's goal had been to please others, then C1 is the rational choice. But it's not the goal; let's keep the goal fixed as stated above, namely, the agent insists that reliable transportation his goal. Given the agent's social nature, however, the rational choice, C2, is hard to make, and the agent experiences strong internal pressure to choose C1. But why is this? No one is "forcing" the agent toward a C1 decision. How is the agent's decision being compromised? One way we might think of decisional compromise is that the agent's rational "self", the agent's practical reasoning, pulls toward a C2 choice, but the agent has a competing social "self" that

pulls toward a C1 choice. A strong social self, then, compromises the agent's rational self, and the more "social" the agent is in personality the harder a rational choice becomes.

Even if an agent in this situation ends up making the rational choice, the agent's social nature is a psychological factor that has made it all-the-harder to be rational. It is easy to see how agents could end up giving in to various social pressures and make a bad decision with C1, given the original goal of reliable transportation. The idea is not that society is openly forcing, coercing, or threatening the agent; that would be intellectually easy to counter. Rather something deeper is going on: the agent has internalized certain cultural values and social norms that now interfere with making a rational choice. It would not help for the agent to choose C3, for the rational choice C2 would still have been missed. The agent would be better off, in our hypothetical example, not caring what other people think, or being ignorant of what choice is socially acceptable. (It is another issue, however, if the agent would be better off over-all not caring about what society thinks of her.)

(b) The **practical fallacy of decisional habit**.

Social habits can become deeply established in our lives. The interesting and troubling thing about social habits is that they often do not appear to be habits to those who have them, they appear simply as the right thing to do, the normal, proper, and natural way to think, talk, or behave. Our social nature is partly made up of such habits. There are patterns of conformity that do not seem to be conformity, but just the normal way of doing things. There are expectations of peers and social authorities that do not appear to be pressures put on us by them, or that we put on ourselves on their behalf, rather we simply fulfill them in the normal course of our day. And there are patterns of conventional wisdom and group-think that we automatically put into our words, often as clichés, and these come out of our mouths so effortlessly and naturally that no thinking seems to be needed.

In one way, our social habits are very efficient fall-backs when we are at a loss for words, or at a loss about what to do. They are almost guaranteed not to get us into trouble with our social peers and groups, and in fact connect us to people in our social setting all-the-stronger, for they are part of our social nature

that others expect to see in us and find comfortably familiar. But social habits can make for irrational choices. Take the case of an agent who has become used to using foul words; her language is full of such words. Suppose that she forms a goal of cleaning up her speech, perhaps she has a new job in which such language is not appropriate. To achieve her goal, she decides to be mindful of her speech, to censor herself using foul language and substitute less “colorful” vocabulary. But this agent belongs to social groups in which she habitually talks in foul language. Without having given up her goal of cleaning up her language, she nevertheless lets herself slip into using foul language when she is socializing. Given her goal, this is an irrational choice caused by the agent’s social habit of language use.

This example could easily be multiplied to other social behaviors besides language use, dress for example. The internal force of social habit, in keeping with our social nature, can “pull” a decision away from the option that should be chosen toward one that is an irrational choice.

3.3 Practical fallacies arising from decisional conflict

The second category of aspects of human nature that can make for irrational choice involves the experience of **decisional conflict**. We will look at three kinds of decision conflicts that can interfere with practical reasoning and lead to poor decisions.

(a) Approach-avoidance conflict: An agent has an option whose outcome maximizes utility, but the option has a non-outcome consequence the agent especially fears. The agent, then, is at once attracted to and repelled by the option. The agent experiences a “go – don’t go” decisional conflict. For example, an agent whose goal is to feel more relaxed wants another drink (which will relax him more), but fears becoming drunk and acting offensively. Or an agent whose goal is to satisfy a smoking addiction desires to light up a cigarette (thereby gaining the goal) but has become very worried about the health consequences after a recent physical examination. Or an agent whose goal is companionship would like to meet that attractive person at the party, but deeply fears rejection. In each such case, the

attractiveness of the option's outcome is equal to the repulsion of the potential bad consequences, and the agent feels caught between choosing the rational option and not choosing it.

(b) Approach-approach conflict: The agent is confronted with two equally good options which, however, are mutually exclusive. Given the agent's goal, each option has an outcome that equally achieves it (both are equally rational choices), but opting for one means losing the outcome of the other. (This case is a double approach-avoidance conflict, with emphasis on the approach side of the decision). The agent wavers back and forth between options, the attractiveness of each equally pulling the agent toward a decision in its favor and the loss of each equally pulling the agent away from deciding on the other. This conflict has been famously modeled as "Buridan's Ass." The medieval philosopher Jean Buridan envisioned a hungry donkey caught exactly between two equal bails of hay. Not having a reason to choose one way over the other, and desiring not to give up a bail, the conflicted donkey starved to death. Take the more familiar case of a hungry agent whose goal is a nice meal at a good restaurant. The waiter offers two specials of the day at the same reduced price: a delicious seafood dish, and a steak special. The agent equally loves seafood and steak, and tonight is in an equal mood for each special, but can only eat one. The waiter is waiting for a decision, but the agent can't seem to decide, each option being equally appealing and neither showing a flaw that would break the tie. The waiter leaves and returns later, but still no decision. The experience of conflict mounts as the agent feels pressure to make a decision, but can't find a reason to choose one special instead of the other and doesn't want to let one go in favor of the other. The more the agent looks each special over, and the more he tries to discover which one appeals to him more, the more he is caught in an approach-approach conflict.

(c) Avoidance-avoidance conflict: The agent is forced to decide between two options each of which has equally undesirable outcomes (i.e., each outcome has equal disutility, given the goal). These kinds of decision conflicts are often called "practical dilemmas" or "hard choices." Harm will be done no matter what is decided, and yet there is no way to escape a decision, given the goal. There is no maximin "lesser of two evils" in this conflict, for the options are equal in the consequences the agent desires to avoid. (This is a double approach-avoidance conflict, with emphasis on the avoidance side of the

decision). Here is a dramatic example. Imagine a war in which a commander must capture a certain territory or all his troops will be killed. His goal is to keep all his troops alive. There are only two options presented to him by the analysis team. He can capture the territory by a bombing effort that will take about one week but an estimated half his troops will be killed. He can capture the territory by a shelling effort that will likewise take about one week but again an estimated half his troops will be killed. The commander can't bring himself to decide which strategy to go with, each equally revolts him in the loss of life. Another example: the novel "Sophie's Choice" by William Styron describes a particularly horrifying case. In a Nazi concentration camp a mother is forced to choose which of her two young children will be taken away to be killed. She can't do it, but then is told that if she does not choose one, both will be killed; she must choose but still can't until the very last moment. For a less dramatic example, suppose an agent has an important job interview to get to and is running late due to traffic. His goal is to get the job. He can't receive another speeding ticket or his driver's license will be revoked, and he can't get to the interview late or he won't get the job. If the agent speeds he risks a speeding ticket, but if he drives at the speed limit he risks being late for the interview. The agent is caught in an avoidance-avoidance conflict and experiences this decisional conflict between speeding or obeying the speed limit. As a final example, take an agent whose goal is to respond to a good friend. There are two options open to the agent: tell the truth or tell a lie. Telling the truth will deeply hurt the friend, but telling a lie will violate the agent's moral values. Each consequence is equally repugnant to the agent, yet one or the other must be done given the goal. In avoidance-avoidance cases like these, each option equally repels the agent, for the agent desires to avoid harmful consequences. The need for a decision, however, cannot be avoided. Having to choose to do something that the agent knows will result in harm makes the avoidance-avoidance an especially disturbing conflict for an agent to experience.

How do decisional conflicts lead to bad decisions? Notice that each kind of conflict produces a state of **indecision** in the agent, and yet the agent either must or desires to make a decision (given the goal). Thus, there are two layers of conflict: the base level is a tug between approach and avoidance, or between approach and approach, or between avoidance and avoidance. Each of these three results in a state of indecision. And then on a second level this state of indecision conflicts with the agent's need or

desire (or, in Sophie's case, pressure) to make a decision. A decision, therefore, will now accomplish two things: it will move the agent toward (or away from) a goal (on the base level), and it will end the agent's state of indecision (on the second level). Relief from indecision now becomes a complicating part of the original conflict. The classic responses to decisional conflict show how escape from indecision becomes an added burden to making a good decision. Let's review these patterns of bad decision making.

- One response is for the agent to minimize practical reasoning and make a quick decision just to get it over with. This is the **fallacy of making rash or hasty or snap decisions**.
- Another response is to have someone else do the practical reasoning and come to a decision for the agent. The agent who can't decide between seafood or steak for dinner might ask the waiter, "Which do you recommend?", or might tell a friend, "You decide for me." This is the **fallacy of decisional passing**, a way of avoiding practical reasoning and the responsibility of making a choice.
- Another response to decisional conflict and indecision is to fall back on habitual or familiar ways of choosing. The **fallacy of decisional habit** has already been covered above from the point of view of our social nature.
- Yet another response is to try to put off a decision until the last moment, or try to defer it until the opportunity or need for a decision is past. This is the **fallacy of procrastination**, which often takes the pattern of the agent allowing other tasks to take undeserved priority in the agent's life, and then using "being overextended" as the excuse to put off the decision.
- Another response to decisional conflict and indecision is to leave the decision up to chance, the **fallacy of making random decisions**. The agent might flip a coin or resort to some other chance event that decides the issue. The agent at the party who fears rejection might say to herself, "I'll flip a coin and if it's heads I will not introduce myself to that person, but if it's tails I'll push myself to do it."
- Another classic response is to complicate the decision to the point of confusion. The agent attends to and begins to worry over every little detail of the decision, multiplying irrelevant small factors to the point where the decision becomes too complex and confusing to make. This is sometimes referred to as the state of hyper-vigilance. The agent lets himself off the hook, in a certain sense, through hyper-vigilance, because now the decision appears to be too complicated and confusing to make, or to make well. This is the **fallacy of decision complication**. Becoming

overwhelmed by decision complications is used as an excuse to avoid a decision, or used as an excuse for a bad decision.

What do these familiar responses have in common? They serve to reduce the decisional conflict and pressure experienced by the agent who is in an uncomfortable state of indecision. But relief from decisional conflict and escape from indecision was not the agent's original goal. Decisional conflict interferes with rational choice because the agent has not given up the original goal, yet is now motivated and prone to make a decision whose outcome is conflict reduction and indecision removal, something unrelated to the goal. The agent caught between risking a speeding ticket and risking being late for the job interview has not given up the goal of landing a needed job. But this agent might decide to deal with the decisional conflict and resulting indecision by hastily turning around and going home or rashly taking an unfamiliar road in the wild hope that it is a short cut to the interview. Given the goal, these are not rational choices.

3.4 Practical fallacies due to personality weakness

The third category of factors in human nature that can make for irrational choice involves certain weaknesses in our personalities. Most, if not all, people have weaknesses. There might be a particular item of food that a person should not eat but can't resist. There might be a particular item for sale that a person should not buy but can't stop from purchasing. There might be a certain person that you should stay away from but can't bring yourself to stay away from no matter how hard you seem to try. There are two widely studied weaknesses that interfere with making rational choices: (a) the problem of sunk costs, and (b) the problem of binding the will.

(a) The practical fallacy of sunk costs.

Sometimes an agent chooses a course of action that yields its outcome quickly; in considering the option, the agent foresees the need for only a small effort to result in the intended outcome. But many decisions are not like this; an option might involve a course-of-action that would require time and effort on the agent's part if the outcome is to happen. In such cases, it is not enough for the agent just to make the

decision; the agent must also “stay the course” or “see it through”. Staying the course, once a decision is made, might mean temporary setbacks, overcoming hurdles, and readjusting plans and strategies on the part of the agent. A decision that needs to be sustained over a period of time will typically require the agent to invest herself and things she values in the decision. These are the agent’s **costs**. (See the explanation of costs within the definition of outcome in Chapter 3, section 3.1.2)

Some costs might be recoverable once the goal is achieved. If an agent’s goal is to make money, for example, and if the option the agent chooses as the best way to reach the goal requires the agent to spend money, these financial costs can often be recovered once the goal is achieved. But some costs of staying the course and seeing a decision through to its outcome can’t be recovered; they are “sunk” into the decision and forever gone. **Sunk costs** are all those things of value to an agent that have been invested in a decision and which cannot be recovered. These can be things like money, time, effort, the fact of having tried, energy, talent, materials, wear and tear on the agent’s body, or wear and tear on the agent’s tools or equipment. They can also be less tangible but no less important things like the agent’s reputation, pride, sense of self, and other aspects of the agent’s ego. There is one sunk cost connected with every decision, namely, the next best option the agent could have chosen but didn’t. The chance to do something else at that given time is forever gone with each action an agent decides to do.

How do sunk costs lead to bad decisions? The more costs an agent has sunk into seeing a decision through, the stronger the agent must be to opt out of the decision if it is going wrong. Some decisions, due to various changes that may have taken place in an agent’s life, are no longer decisions the agent should be in. The rational thing to do when a decision becomes bad is for the agent to opt out. But, if the agent has invested herself and her valuables in the decision and these are now sunk costs, they tend to trap an agent into continuing with the decision and into staying the course. This irrational choice is the fallacy of sunk costs. The more sunk costs, the harder it is to opt out when it becomes rational to do so. Agents have weaknesses about sunk costs, and often cannot find the strength to opt out. Here are some familiar examples involving money, and then we’ll point to some non-monetary cases.

A person who has invested a lot of money keeping his car repaired is now faced with yet another needed repair. The car is old and in poor condition. It will cost another, say \$500, to get it repaired. In the last few months the agent has already spent over \$2000 in repair bills. The car is not worth \$2500, and the agent knows this. If offered the same old car in the same poor condition to buy for \$500, the agent would make a good decision and turn the offer down, seeing the car for what it is – nothing but a lot of trouble and needing a lot of work to keep it going. But the thought of having already sunk \$2000 in the car makes it too hard for the agent to opt out and get rid of the car, and he decides to sink another \$500 into repair costs.

Take another example, the case of a homeowner who must drill for water because the old shallow well has gone dry. He hires a drilling crew that charges \$10 per foot. After several days they have drilled 500 feet down and still no water = \$5000 sunk costs. The crew offers the homeowner two options: abandon the site and start drilling on a different part of the property, or keep drilling another 100 feet below which there is sure to be no water. They offer their advice to the homeowner: in their experience there is only a 5% chance of finding water in the next 100 feet where they are now drilling, and a 70% chance of finding water within 500 feet down if they try another site. They recommend opting out and starting over. To the homeowner, that's \$5000 down the drain. If, from the start, a drilling crew had offered to drill for water 100 feet down at \$10 per foot in a spot that had only a 5% chance of striking water, the homeowner would have correctly said no. And yet after sinking \$5000 already in one spot, the homeowner is too weak to opt out and has the crew go the next 100 feet, "just to be sure."

Here are some non-monetary examples of the problem that sunk costs creates for rational choice. Companies and nations find themselves irrationally staying the course in a decision they should opt out of, because of such sunk costs as reputation, image, pride, and the like. Often people remain in relationships that are clearly not good for them to be in, and yet can't opt out because they have already put so much of themselves into making the relationship work and don't want to admit to "failure."

The problem of sunk costs is the weakness that agents show in giving up a decision that is no longer rational for the agent to be in – a weakness that is due to costs the agent has already sunk into staying the course. It is not rational for an agent to bind himself to a decision with sunk costs to the point beyond which the agent is not strong enough to opt out and “cut one’s losses” should the decision go bad. Too much commitment to a decision is sometimes not a good thing, for if the decision should no longer be seen through by the agent, the agent might not be strong enough to overcome the power of the commitment in the form of sunk costs, be they material or spiritual.

(b) The practical fallacy of binding the will.

The problem of binding the will is the flip-side of the problem of sunk costs. It is weakness at the other end. Sometimes an agent makes a decision and is not strong enough to stay the course and invest what it takes to see the decision through. The agent wants to reach a goal, and knows the best option to achieve it. The agent further realizes that the option will involve setbacks, hurdles, difficulties, temporary failures, temptations to give up, sunk costs, and in general will take strength, commitment, and perseverance to stick with the decision. The agent feels weak facing such a daunting decision, and lets the option go by, or in the case of a decision already made, opts out as soon as things become challenging. The agent might not wish to appear to be a failure either in the eyes of others or in the agent’s own eyes, and so will decide not to take a course of action, or will decide to give up a course of action in which there is the threat of failure. And yet, given the agent’s goal, the decision is one the agent should see through; to choose otherwise would be an irrational choice: the fallacy of binding the will (that is, failing to bind one’s will to a good decision the agent should stick with). The classic way to present such bad decisions both to others and to oneself is the “sour grapes” explanation. The sour grapes reaction is the attempt to devalue the goal (falsely) to the point where passing on, or opting out of, a challenging decision looks like the “right” thing to do.

Binding the will is a decisional problem for two types of agent: those who tend to give up too easily as a general character trait, and those who foresee in a decision a situation in which they know themselves to have a particular weakness. An example of the former agent would be a person who finds nothing of

interest in life, nothing seems worth the effort, and nothing is of real value for this person. Such a person would, quite naturally, not be committed enough to make the effort required to see a decision through, if the decision was one that required the agent to bind the will. An example of the latter agent would be someone with a specific addiction, say smoking. The person decides to quit, but knows that she will be especially weak in keeping to this decision and will be tempted to override it with a decision to smoke, say, in situations in which friends are smoking.

For either type of agent, knowing our weaknesses allows us to take measures to help “bind our will” to the option that is the rational choice. There are various incentives, disincentives, and mechanisms that an agent can set up that make it difficult, and perhaps even impossible, to opt out of a decision due to lack of strength to see it through. Take the case of someone addicted to smoking whose goal is to quit smoking; let’s suppose that there are major health concerns connected with smoking that now make quitting desirable. This agent decides to stop smoking today. However, the agent realizes that he is not strong enough to overcome his addiction to cigarettes. Based on past experience the agent knows that he will not be able to resist the temptation to smoke, especially in certain situations: when socializing with other people who smoke, when drinking alcohol, after a meal, and when experiencing anxiety.

What might this agent do that would bind him to his decision to quit smoking especially in those situations where he is most tempted to opt out? The agent might set up a system of better and better rewards for each day that goes by without a cigarette. The harder it gets to resist a cigarette, the bigger the reward the agent forces himself to give up for smoking a cigarette. The agent might set up a system of harsher and harsher punishments, so that the harder it gets to avoid smoking, the bigger the punishment the agent can avoid by staying the course. The agent might make a promise to himself to avoid all those situations – socializing with others who smoke, or drinking alcohol, or eating out – in which he expects to become especially weak, until all desire to smoke is gone. If he gives into the urge to smoke in one of these especially weakening situations, he will perhaps have an excuse for smoking, but will have no excuse for breaking the promise to himself to avoid such tempting situations and now has to own up to being the kind of person who can’t trust himself. An agent might request help from friends to stay the

course. If the decision to quit smoking is serious enough, the agent might even place himself in a situation in which it is impossible to smoke, perhaps confinement of some sort, as is often the case with more serious addictions. All such possibilities have the affect of binding an agent to a decision when the agent anticipates a weakness that will tempt the agent not to stay the course.

The metaphor of “binding” comes from a famous episode in Homer’s “Ulysses”. Ulysses had to steer his ship close to the land of the Sirens. Men, it was told, could not resist their song, and as a result sailors would steer their ships into the rocks and all would die. Ulysses desired to hear such irresistible singing but did not want to give into it and imperil his sailors. So he had their ears blocked with wax, had himself bound to the mast, and aimed his ship close to the land of the Sirens. Once he heard their singing, he could not resist and wanted to go closer. He yelled for his men to sail toward the singing, but now they could not obey. He struggled with his ropes, but to no avail. He could no longer reverse his decision, exactly what his earlier stronger self planned for his now weaker self. His bonds forced him to stay the course, his rational choice, when his weakness would have allowed him to choose irrationally. The lesson is that there are certain decisions in which being bound and not being able to reverse or drop a decision is a strength; they are rational choices and being able or having the freedom to opt out is a potential weakness for it allows giving in to temptation. The very opposite is the case in the problem of sunk costs.

3.5 Practical fallacies due to value distortion

The fourth category of features about human nature that can make for irrational choice involves confusions and distortions in what we desire and value. Recall that practical reasoning has two basic components – the agent’s beliefs and the agent’s desires – that result in a decision about what to do. In deliberating about what to do, an agent is continually valuing (desiring) the goal, and is evaluating (forming beliefs about) the best means to achieve it. Someone might object at this point: how is it possible for an agent to confuse or distort what the agent values? An agent’s desires are its own business; on what basis can we, or anyone for that matter, ever tell that an agent’s desires are distorted?

This is an excellent point! It is quite right that for any *one* decision an agent makes, we would have no basis for saying that it is an irrational choice due to value distortion. For this category of practical fallacies, then, we must use two decisions and compare one with the other. The principle of practical reasoning we will use here is this: if two decision situations are equal with respect to all the relevant information required to make a decision, then an agent having this information should make the same decision in each of these two cases. Put more briefly: decisionally equal situations should result in the same rational choice. Value distortion fallacies are decisions that violate this principle. The two areas of value distortion we'll look into are: (a) the problem of relative value, and (b) the problem of discounting the future. To see how these two types of value distortion can lead to patterns of irrational choice, and to appreciate some common mistakes we make about values, it will be helpful first to set up a framework of general ideas about value and value judgments.

The general study of value is a part of philosophy called axiology. The ideas that we will use to form our value framework come from axiology and not from the study of practical reasoning. As noted in Chapter 1, one kind of value that is central to practical reasoning is subjective value, and is thought of as strength or degree of desire. A strong desire for something means that the thing desired has a lot of value for the person desiring it; the person sees the thing as very valuable. Another way to describe subjective value is degree of importance. If a person gives importance to something, then it has value for that person. Recall the order of priority in the case of subjective value (see Chapter 1). It is not that a person desires something *because* it has subjective value for the person. Rather, it is the other way around: something has subjective value for a person *because* the person desires it. Value results from desire, not the other way around.

Typically, we do not keep our values to ourselves. We express them to other people in the form of value judgments. A value judgment is a claim that something has value, or that something has a certain amount of value, or that something has more or less value than something else. Sometimes, our value judgment is "negative". Corresponding to value there is disvalue or negative value, and we put this into a judgment

when we say that something is bad, or undesirable, or unimportant, or is more trouble than something else. We will use “value judgment” to include both judgments of positive and negative value.

We want to distinguish value judgments from a very different kind of judgment called an “evaluation”. To evaluate something is to rate it, to measure it against some standard, or to estimate how well it meets a criterion. Evaluation is an important part of practical reasoning; we value our goals, but we evaluate our options and outcomes. Evaluations often go by the name “normative judgments”. The difference between value judgments and evaluations is often not clear when people express themselves. Let’s take some examples. Suppose someone says, “Coke is better than Pepsi”. If this is a value judgment than it means, “I like Coke more than I like Pepsi.” or “More people desire Coke than desire Pepsi.” But if this is an evaluation, then it means, “Coke more than Pepsi meets such-and-such standards for being a soft drink.” Suppose someone says, “Toyotas are good cars.” What could this mean? If it is a value judgment, then it means, “I prefer Toyota cars over most other cars.”, or “People like Toyota cars a lot.” But if it is a normative claim, then it means, “Toyotas rate highly given such-and-such automobile criteria (e.g. braking distance, comfort, fuel efficiency, safety, frequency of repair, resale value, etc.)” In the case of value judgments it is a contradiction for a person to say things such as, “X is good, but I (or people) don’t like X”, or “X is better than Y, but I (or people) desire Y more than X”, or “X is bad, but I (or people) like it anyway”. These are outright contradictions because the left side value judgments already mean strengths of desire. In the case of evaluations, however, it is perfectly fine for a person to claim, “X is good, but I (or people) don’t like X”, or “X is better than Y, but I (or people) like Y better”, or “X is bad, but I (or people) desire it anyway”. These are not contradictions because the left side evaluations have meanings that are independent of degrees of liking or disliking.

Keeping value judgments distinct from evaluations helps us avoid some common mistakes. Sometimes we hear people claim that evaluations are nothing but personal tastes. Clearly, this is false. Evaluations measure distance from a set of norms, standards, or criteria. Just as we are able to measure a distance of, say, three miles only by applying a standard of measurement, we are also able to measure the performance of, say, a car only by applying performance standards for automobiles. The same goes for

movies, music, painting, poetry, food, clothing, animals, and everything else. To evaluate such things means judging by a set of norms how close or far they are from the ideal the norms represent. There is nothing necessarily involved about anyone's personal taste. Value judgments, however, have everything to do with a people's personal tastes. These, remember, are an expression of how much a person desires something and desires are closely related to personal tastes.

Another error is to confuse both evaluations and value judgments with "mere opinion." Often a person's value judgment, or an evaluation they make, is said to be "just their opinion." Any judgment a person makes might be "just their opinion," even factual statements. For example, someone might say, "That tree is an oak, but that's just my opinion." This is a factual claim, not a value judgment or an evaluation. To add the phrase "but this is just my opinion" means that the person admits they could be wrong, and that they do not have any evidence or argument to back up their position. The position might be about anything – values, facts, peoples, places,.... . Just because a judgment is about value, or a claim is an evaluation, does not by itself mean that there is nothing more to it than opinion. There may be very good evidence that a movie, say, fails to meet certain cinematography standards and in this respect is a bad movie. Likewise, there may be a vast amount of evidence that a certain model of car meets braking criteria to a high degree and in this respect is judged to be outstanding. This is also the case with value judgments. There may be a great deal of evidence that an object is highly valued by many people. To call it valuable would, in this case, not be a matter of "just someone's opinion." The obvious example is money; it is almost universally desired and so if you said that money is very (subjectively) valuable it would be wrong for someone to respond that "this is just your opinion."

A third frequently found mistake is to equate value judgments, as well as evaluations, with "personal bias" or "prejudice". Sometimes the term "subjective" is used to indicate the presence of psychological associations, assumptions, or biases that an individual, or group, links with a statement. If someone claims that a certain part of town is the best place to live, and then we find out that she lives in that part of town, we might call the claim "subjective" meaning that the person is biased in favor of her neighborhood. If someone asserts that his country is the best on earth, we might call this "subjective" meaning that the

person had been indoctrinated into believing this. But notice that “subjective” in this sense can also be applied to any claim, not just to value judgments and evaluations. I might say to you that Dr. Smith is a heart surgeon, and you might automatically assume that Dr. Smith is a man. You would be wrong. You would be adding a “subjective” assumption to my purely factual statement. If I said to you that the children’s nanny had three references from former employees, you would again probably “subjectively” assume that the nanny is a woman. Again, you would be wrong. But you would be wrong about a statement that has nothing to do with values; it is a purely factual claim. Any belief at all, a factual claim, a value judgment, or an evaluation might be biased, prejudiced, and linked to all sorts of assumptions. But no statement need be. Because a judgment is about value, or is an evaluation, does not automatically mean that it is “subjective” in the sense of biased. Evaluating an Olympic athlete as deserving the gold medal, or claiming that you value (=desire) education a great deal, need have no bias, prejudice or other “subjective” association connected with it at all.

Using the distinctions between value and disvalue, and between value judgments and evaluations, let’s now return to the topic of practical fallacies and look at the two kinds confusions or distortions of value that interfere with rational choice. As mentioned above these are (a) the problem of relative value and (b) discounting the future. Both of these decisional problems involve value judgments, but not evaluations. Within value judgments, they include value as well as disvalue judgments.

(a) The problem of relative value.

A person might value something independently of anything else, purely on its own. A couple might desire to have a child, or an item might have great sentimental value for someone. These would be examples of intrinsic (subjective) value or importance, intrinsic in the sense that the value something or someone has for a person is fixed and stable in the thing desired or loved; its value depends only on the person’s degree of desire and is independent of variations that happen elsewhere. In contrast to intrinsic subjective value, some things are valued relative to a situation or context, and the amount of value such things have for people depend upon variations in the situation or context. This is called “relative value”: if item X has value (v) for someone, and (v) is variable relative to something else, then (v) is X’s relative

value. For example, a given amount of money may be worth so much in one part of the country, but in another part of the country, say in a large city, that same amount may be worth significantly less. An item of clothing may be very desirable in one climate but not desired at all in another climate. In the first case we have subjective value varying with locations, and in the second example it varies with climates.

An important kind of variation happens in the amount of value something is estimated or perceived to have relative to amounts or degrees of itself. Let's take money again. Relative to a base of little money, \$100, say, looks very desirable to someone. But relative to a base of millions, that same \$100 holds little or no attraction for the person. An item of clothing might be desirable to a person who does not own anything like it, but if the same person already had a closet full of such items, it might have little value, and perhaps even have disvalue for the person. This form of relative value is referred to as the principle of proportionality. It states that for some kinds of relative values, the amount of it that an item is perceived to have or is given varies in proportion to the base amount of it. The old saying, "absence makes the heart grow fonder" is a case in point. The less presence there is of a person we are fond of, the more we desire (value) the person's presence. Note that the principle does not specify how the proportion varies. It might be the case that as the base amount of value grows larger, the value the next portion of it has for a person becomes smaller. This is the case with the examples of money and item of clothing just given. There is a well-known case of the principle of proportionality in economic theory called the law of diminishing marginal utility (=value) that captures this idea. This principle states that the more a person has of something the person values, the less value the person perceives the next minimal additional unit of it to have.

There are other ways the proportion could vary, however. Suppose a college was expanding and needed land. More and more land is bought over the years. But suppose that right in the middle of the college's ever-increasing land holdings a farmer still owns fields and refuses to sell any land to the college. In this case, the more land the college owns around the farmer's fields, the more desirable those fields become to the college. The college might offer the farmer a price to sell the land that is far higher than the land

would sell for on the open real estate market (minus the college as a potential buyer). This variation of proportional relative value is different from that described in the law of diminishing marginal utility.

How does relative value create a problem for rational choice? It does this by distorting an agent's perception of value in a decision situation, a distortion that leads to a bad decision. This is the **practical fallacy of relative value distortion**. Let's look at familiar examples. Take two stores that are located a short distance from each other. Shoppers go to one store and buy items they need that are on sale there, and then go to the other store to buy the other items they need that are on sale in this store. Take a typical shopper who decides to go the distance for an item on sale for \$1.49 in the other store that costs \$1.99 in the store he is in. He has gone the distance and now he is in this other store and has bought the item on sale for \$1.49; he saves \$.50. Next he is ready to make a big purchase costing \$99.99 and remembers that the other store has it on sale for \$99.49. But instead of going back to the other store, however, our shopper now decides it's not worth it. Do you see the problem? Notice that these two decisions, traveling to save \$.50 in one case and not traveling to save \$.50 in the other case, are inconsistent. One of these is a bad decision. (Leave aside issues like being tired, having to carry more things, and other possible explanations of the inconsistency.) The problem is that the \$.50 has not changed value, so if it was worth the effort in one case it should be worth the equal effort in the other case. If it's not worth it for the agent in one case, it should not be worth it in the other. The two decision situations are exactly the same, invariant, in all relevant aspects. What was different – the price of the items at \$1.99 and at \$99.99 – is not relevant, for the decision is made on the basis of the amount saved and that was exactly the same in each situation. Practical reason requires that an agent should make consistent choices in decisionally identical situations. What accounts for the irrational choice (whichever one it was)? It can only be that the agent perceived saving \$.50 in one case to be desirable (valuable) in proportion to \$1.99, but saving the same \$.50 to be of no desire (value) in proportion to \$99.99. The problem of relative value causes the fallacy of relative value distortion.

There are many other familiar examples. A person selling a house for \$275,000 is unwilling to accept an offer of \$260,000. But if the house had been for sale for \$750,000, the same seller would be willing to let

it go for \$735,000. Why? In one case \$15,000 looks like a lot, but in the other case the same amount of money seems small. Imagine that a person bidding at an auction has just bid \$25 and is outbid by someone who bids \$30. The person decides not to top the \$30 bid. But if the same person bids \$250 on an item she desires no more than the item she let go for \$30, and if now she is outbid by someone who bids \$255 for it, she might find it easy make an inconsistent decision and bid, say, \$256. Why? \$5 has distorted value relative to \$25 in comparison to \$250.

In each case in these examples, one decision is a bad decision. It does not matter which was the bad one, because these examples only call for consistent decisions in equal situations. That they are not perceived by the agent to be equal situations is exactly the distortion that makes relative value a problem for rational choice; the frame of reference relevant for each decision has really not changed. One final example, designed and studied by the psychologists Daniel Kahneman and Amos Tversky [Plous, pp. 74-75]. Subjects were asked to imagine that they were on their way to the theater to see a play. Tickets cost \$10. At the ticket window you realize that you have lost \$10. Would you still buy a ticket to see the play? Most of the subjects (88%) answered that they would. The same subjects were asked about the same decision situation, with an irrelevant twist put in. Imagine that you already bought the \$10 ticket, and at the theater realize that you have lost it. Would you spend \$10 to buy another ticket? Only 46% of the subjects said they would. Relative to \$10 already spent on the lost ticket, seeing the play appears to cost \$20, whereas relative to the lost \$10 the play appears to cost only \$10. But decisionally these two situations are identical, and call for consistent choices. Switching the decision from buying a ticket in one case to not buying a ticket in the other case means that one of these decisions is irrational. Value distortion due to relative value seems to be behind such inconsistent decisions.

Another form of relative value distortion, related to the last example, is called **loss aversion**. Here is a little imaginary experiment you can test yourself with that illustrates loss aversion. In which case would you feel worse? (1) While walking you suddenly see a \$100 bill, but before you can get to it (finders-keepers!) someone beats you to it. You're out \$100. (2) You had a \$100 bill in your pocket, but when you go to get it you realize that you lost it. You are out \$100. Now if you are like most people, you'll feel a lot

worse in the second case than you will in the first (at least you'll respond that you would feel worse in case (2)). But why? In each case you are in the same situation: out \$100. It seems that people generally experience greater hurt *loosing* something of value that they possess, than they experience *failing to gain* the very same thing. Why should this be? In both cases, they are without the item of value. Perhaps people tend to value the possession of a thing that they value, and add this extra "possession-value" to the thing's value, whereas they don't (can't) do this with a thing they value but don't possess. Whatever the explanation, loss aversion means that people will tend to "hang onto" what they have, and to "over-estimate" its value, compared to the same or similar things they don't have. Looked at the other way around, loss aversion means that people tend "under-estimate" the value of things they don't possess compared to the same or similar things they do possess. Somewhere there is value distortion going on. For a typical concrete example of loss aversion, ask a person who has bought a new car how he likes it. He'll no doubt answer that it's a great car. But if the same person's neighbor was the one who bought that new car and had loaned it to the person for the same amount of driving time, and we asked how much he (the non-owner) likes the (neighbor's) new car, we can be pretty sure that it won't be as much as he would have liked it as the owner. Here is another example: ask a person what she thinks is a fair price to ask for something she values and owns, if she were going to sell it. Compare her answer with the amount she gives, if at another time you ask her what she considers a fair price to be for that same item if she were going to buy it. Loss aversion predicts that the first figure she gives will be higher than the second figure.

You can imagine the potential loss aversion has to influence decisions. Say a friend approaches you for advice: should she trade her car in for a newer model? You know her situation and her car pretty well, and offer your honest advice: yes, she should get rid of it and go for the newer model. A week later, you get your car back from yet another repair and toy with the idea of trading it in for a newer model. You decide to keep it; you think, it's still got "a lot of life" to it yet. Your friend is surprised at your decision; she points out, correctly you realize, that your two circumstances are totally similar as to money, age and condition of cars, driving needs, etc. These decisions should have come out the same; you both should

decide either to keep the present car or get rid of it, for similar decision situations call for similar decisions. One of these decisions is bad: the one given as advice to your friend to sell it, or your own to keep it.

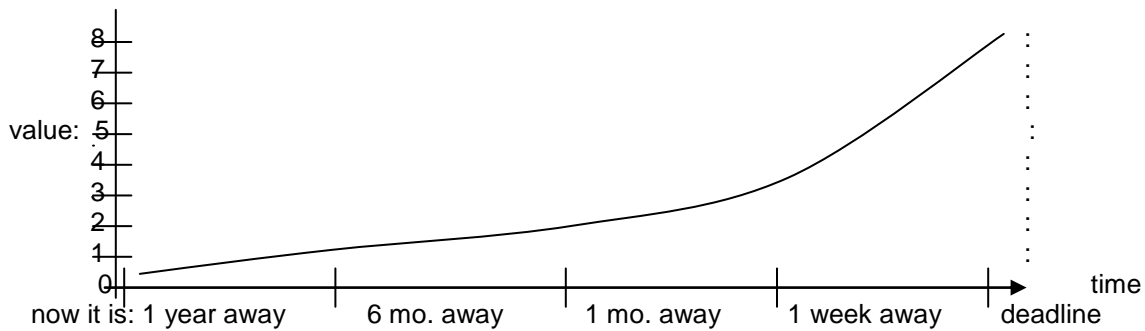
The practical fallacy of loss aversion can be especially troubling in a situation of bargaining and negotiations, as you have seen in Chapter 13. It is typically the case that agents must make concessions in order to settle a conflict by bargaining. Concessions are not only required in the area of what an agent hopes to gain, they are often expected in the (painful) area of what an agent has already gained and counts as owned. If an agent sees a concession as a failure to gain what the agent had hoped to gain as her part of the bargain, it might be easier to make the concession than if the agent sees that same concession as a loss of what has already been achieved. In deciding whether or not to concede something that the agent values (say, a certain percentage in salary and work benefits in re-negotiating a labor contract, or certain holiday visiting rights with children in a divorce conflict), loss aversion might even make it impossible for an agent to concede anything, and the entire bargaining process might collapse. Here is an example. Take two countries that have been at war with each other and are now trying to negotiate peace. Let's assume in this example that continuing war is a bad decision, and negotiating a peace treaty is the right course of action. One side has lost territory to the "enemy" and wants its traditional land back. The other side believes it has won this territory from its "enemy," at the cost of many lives of its "brave soldiers." What happens if we add loss aversion to each agent in this not-so-hypothetical territory conflict? It becomes much more likely that the peace negotiations break down, because now the hard decision that each side faces to concede all or part of the disputed territory to the other side becomes next to impossible to make. And so (we imagine) they return to war, each side blaming the other for being "intransigent" and responsible for the failed negotiations.

(b) The problem of discounting the future.

Time and value are two central features of human existence and are related to each other in systematic ways. For example, people would prefer, all other things being equal, that bad experiences (disvalues) were in their past rather than in their future, and good experiences (values) were in their future rather than in their past. In addition, people would prefer that bad experiences were in their distant past or their

distant future rather than in their near past or their near future. But it's the other way around when it comes to good experiences; people tend, all things considered, to prefer that good experiences were in their near past or their near future rather than in their distant past or their distant future. In planning for the future, most people, to the degree they can, arrange to put off bad experiences as far into the future as they can manage and have good experiences take place "the sooner the better." These general human tendencies concerning time and value seem deeply rooted in human nature. Apparent exceptions, such as the person who plans to eat all the unpleasant food first and save dessert until last rather than the other way around, can usually be explained by the relative amounts of good and bad expected to be experienced, and so turn out not to be exceptions. Time (that is: when something is expected or desired to take place) and value (that is: how much a person desires something to take place), then, are strongly interconnected and play important parts in our making or failing to make good choices.

Discounting the future is a value distortion involving future time. Most people have a tendency to diminish the value of something the more into the future it is. A vacation we will enjoy three years in the future is not as important to us now as one we will enjoy in three days time, even though they are identical vacations. Most people care about the well-being of the next generation. But as the generations are projected further and further into the future it is natural to find our concern diminishing to the point where it is hard to give any importance to the well-being of descendants very remote in future time from us now. **Discounting the future**, then, describes the following tendency in human nature: for any given item of value, the more future it is the less value a person tends to assign to it in the present, and the more impending it becomes the more important a person tends to regard it. Here is a representation in the form of a graph. The vertical axis represents degrees of value a person gives something, and the horizontal axis represents periods of time.



Suppose a person starts out 1 year away from something the person desires. Assume that the item has a time limit attached to it, so that it must be done, or experienced, or possessed at a specific time. For example, suppose a person will be starting a new job (or getting married, or undergoing major surgery). Let's call the time at which she must show up for the new job the deadline, here due one year into the future. The person values this new job a lot; let's say a 7 on a scale of 0 to 8 with 8 representing maximum value. But because it's "so far" in the future, a full year away, the person discounts the new job 6.5 value units. In other words, the person gives priority of importance to all things in her life that she now, one year away from starting her new job, desires greater than strength .5. Time passes, and the person is now 6 months away from her deadline. She discounts her new job a little less, for now her deadline is not as far into the future as it was 6 months earlier; at the 6 month point it's valued at strength 1. At this point she still gives more importance to most things in her life, namely, all things she values above strength 1. More time passes, and now she is 1 month away from her deadline and she discounts the job from 7 to 2.5. Soon she is 1 week away from her deadline and now the new job becomes a value 4 in her priorities. By about Wednesday of the last week the job's value shoots up to a 6, now discounted only by value strength 1. We can imagine the agent turning more and more of her attention to the new job, for not many things in her life will at this point in time take priority in her value system. By the end of her week, when her deadline is, say, one day away, she starts to value it the maximum amount she gives it: a 7 out of 8 rating.

This graph and accompanying example are meant to give you the general idea of discounting the future, it is not meant to represent the exact value and time amounts for everyone. However, most of us should be able to see ourselves in this example. The more into the future a good (or bad) thing is, the less we now subjectively value (or disvalue) it. The general point is that discounting the future is a systematic way that the perception of value can be distorted by time and that such discounting represents a tendency that seems deeply rooted in human nature. We should note, moreover, that there are good reasons to discount the future and important benefits result. In fact, we should be relieved that our nature is such that the future is discounted rather than the present. Just imagine what we would be like if human nature were made the other way around and there was a strong tendency to discount value as things become imminent and present, and to inflate their value the more into the future they were. Picture what the graph would look like and think of an accompanying example. Clearly, human life would be much harder, if not impossible, if we tended to discount the present rather than the future. Just imagine a hungry person who desires to eat a future meal but has no desire to eat a meal now; the future meal has subjective value but not the present meal. Of course, there could be and no doubt are people who are exceptions to the rule, people who “live” so much in the future, or what they dream the future will be, that they discount the present. Such people understandably have a hard time coping. The point is that there is basic good sense in the way humans tend to link value and future time. But it can also be the source of trouble and irrationality in certain decision situations.

How does this tendency to discount the future make for irrational choice? Practical reasoning and making decisions necessarily involves time, especially the future. Goals are not in an agent’s past, they are future things an agent presently desires to achieve. An agent’s options are not past courses-of-action the agent has done, they are possible courses-of-action an agent can do in the agent’s immediate future. Practical reasoning is an activity in which an agent’s values and future become linked. It is to be expected, then, that the tendency to discount the future can have a powerful affect on our choices. There are two patterns of irrational choice that commonly occur: (a) the practical fallacy of unintended consequences, and (b) the practical fallacy of procrastination.

(a) The practical fallacy of unintended consequences.

Recall that the outcome of a decision is only that part of the total consequences that achieves or threatens the agent's goal. Among the other non-outcome consequences some might be costly to the agent, and costly in a way that an agent might especially fear. This is the case, for example, in an approach-avoidance conflict. If an option has costly consequences that outweigh the value of the goal, then it is clearly a bad decision for the agent to choose that option, for even if its outcome achieved the entire goal, the costs of the consequences would leave the agent worse off overall. But suppose such costly consequences were far enough in the future that the agent discounted their severity, or were so far in the future that the agent considered it unimportant to look into their possibility. If this were to happen, the agent might well choose that option, for it would now look better to the agent than in fact it is. This is the fallacy of unintended consequences. As the day of reckoning approaches, the agent starts to realize how undesirable the costly consequences of the decision really are, or starts to realize that such possible bad consequences should have been considered much earlier. The agent now regrets the decision, and sees in retrospect how bad (= poorly thought out) it was. The problem of unintended consequences makes the decision an irrational choice, and unintended consequences can be and often are the result of discounting the future. No doubt there are other causes of the problem of unintended consequences, such as ignorance or carelessness in reasoning, but the pattern we are focusing on here is the one that involves a special kind of value distortion – discounting the future.

To illustrate this problem consider this example, an imaginary one that is unrealistically streamlined in order to see how discounting the future can lead to a bad decision. Suppose a nation, mindful of human-caused global climate change, must decide between two long-term energy policies. One policy is to develop a national energy system that uses primarily fossil fuels, and the other plan is a national system that uses primarily nuclear energy. Both energy plans have many attractions, and both have many undesirable consequences, especially concerning pollution hazards. The major difference is that the pollution hazards that will result from using fossil fuels are expected to take place soon after the energy system becomes functional, within say 5 years. But the pollution hazards resulting from nuclear fuel are estimated to take place well into the future, starting, let's suppose, between 200 and 300 years from the

time the system becomes functional. Suppose, further, that such undesirable consequences from using nuclear fuel are much worse than those from fossil fuels for the latter can be off-set more easily by human effort. In other words, if we look at the amount of pollution harm from fossil fuels, pretending we are already 15 years later, they look pretty bad. But if we look at the amount of pollution harm from nuclear fuels, pretending that we are already 200 years later, they look much worse than the harm due in 15 years to using fossil fuels. But to the energy planners of our imaginary nation, who have to make a decision now, 200 - 300 years into the future seems a very long time and they discount the future. The undesirable consequences don't look as bad to them now as they would look to them if they could be transported 200 years into the future. As it is, from their present vantage point they become distorted so that they seem the lesser of the two pollution evils. The nation decides to go nuclear for its long-term energy system. As a result the generation 200 years down the road now faces a huge pollution mess and regrets the decision made 200 years ago, calling it one of the biggest errors in the history of the nation.

(b) The practical fallacy of procrastination.

Procrastination is familiar to many people, and we saw above that there is a tendency for it to happen in decisions involving conflict and indecision. The kind of procrastination that happens in reaction to conflict and indecision is deferring or putting off making a decision. Another type of decisional procrastination happens when an agent sets a goal, discovers the best option for achieving it, but then decides not to act. Recall that deciding not to act is itself a decision; it is not the same thing as indecision (see Chapter 3, section 3.1.3).

How can discounting the future lull an agent into this kind of procrastination? Let's go back to the above graph that relates time and value to each other in a way that represents discounting the future. Take an agent who is now one year away from her deadline. "Deadline", remember, is the term we are using for the time that something the agent values, a goal that has a time limit attached to it, comes due or must be achieved or experienced. Let's suppose that our imaginary agent is very much like ourselves in this important way: she acts when her motivation to act is strong enough. In general, we don't act when our motivation too weak to move us away from whatever other action we happen to be doing. This can be

described using “value” instead of “motivation.” We tend to act when things matter to us, when things become important enough; in other words, when we value something enough, and not before. Let’s, then, suppose that the agent we are picturing, as represented in the graph, will put other things in her life aside and will act to achieve her goal once her desire for the goal reaches, say, level 5. If she valued the goal less than 5, then she would never do anything to achieve it, for her life is such that a level 5 is required for her to stop whatever else she is doing and start acting to gain her goal. If you look at the graph, a level 5 desire (value) happens roughly on the second day into her last week before the deadline (draw a straight horizontal line from value 5 and see the time below where it meets the upwardly curving value line). So, at this point in time, but not before, our imaginary agent gives priority to achieving her goal, which she set for herself a year earlier but decided not to act on, given its low value for most of the year. Our agent puts in five days of activity and now, a full year later, she meets her deadline and achieves her goal.

So far so good. The agent appears perfectly rational in the way that she has juggled her priorities and met her deadline. But what if achieving her goal requires a course-of-action taking more than five days? Suppose her chosen option needs a full week of activity before her deadline in order to achieve her goal, or a month of activity, or three months. In these cases, the agent will have put off acting until it is too late to achieve her goal. If the agent can achieve the goal with five days of activity, there is no problem of procrastination. But if the deadline requires, say, one month of activity, the agent will have put off – procrastinated – acting for over three weeks. This is clearly a bad decision; this is the practical fallacy of procrastination due to discounting the future. The agent will regret having let three weeks slip by, once it is clear that her deadline cannot be met in only five days of activity. As a result of procrastinating her goal is unachieved.

Future-discounting procrastination, then, has these 5 features:

- (i) the agent chooses a course-of-action, option O , that will result in achieving the goal;
- (ii) there is a point of time, T , that is required to start O so that it results in the outcome that achieves the goal;

- (iii) the goal is time-limited, that is: it has a deadline, D , that cannot be extended;
- (iv) the agent starts O at a point of time ($T+n$) later than T ;
- (v) during the time from T to $T+n$ the agent decides not to act and instead does other things. Because D can't be extended further into the future, procrastination is a decision that loses the agent the goal – clearly an irrational choice.

Why would an agent procrastinate? Part, if not all, of the problem involves discounting the future. An agent who acts when he experiences a desire level 5 and above, and who now values a goal with a deadline that is one year away a full 7 will decide and act to achieve it throughout the full year. If the agent only discounts the future 2 points per year, from 7 down to 5, he can again be expected to act early on. But if this agent discounts the future more than 2 value points, so that the future goal is now valued below 5, the option will not be acted on until it's too late. It is reasonable, then, for an agent to discount the future only to the point on his or her value and time graph where these two intersect: (1) the time required to meet the deadline, and (2) the strength of desire an agent must have to stop doing other things in his or her life and start acting to achieve the goal. Any greater discount of the future leads to procrastination, an irrational choice.

3.6 Summary

This ends the topic of fallacies of practical reasoning, patterns irrational choice due to interesting features of human nature. Let's summarize this material, both as a guide for review and to highlight the main ideas.

There are four important aspects of human nature that are closely connected to common ways people fail to make good choices.

- We are by nature social.
- We experience conflict in certain decision situations that leads to indecision.
- We find it hard to walk away from situations in which we have invested ourselves and hard to stay in situations that challenge us.

- We over or under value things depending on their relation to what we already have and to our future time-frame.

It is important to see that these four aspects of human nature are not in any way being condemned as bad. Quite the contrary! If we were not social, the human species surely could not have survived. As for conflict, there are medical researchers who believe that the experience of conflict leads to a good kind of stress that serves to boost the human immune system response. Also, it is well known that the experience of conflict releases hormones that enhance a person's physical endurance. Looking at the third aspect, it is clear that increased commitment to people and situations in which we have invested time, value, and self is a good thing; just think how people would behave if the opposite were the case. As to challenges, they often contain potential harms and threats to our well-being, so it is only natural to feel fear and caution when confronted with challenges. If running from a challenging situation increases your chances of living, while staying and fighting increases your chances of dying, running makes perfect sense unless there are very powerful reasons to stand and fight. Finally, the amount of desire we have for things and the resulting value we see in them from our subjective point of view may be in some sense distorted, but this distortion usually leads to appropriate action or inaction and so on average serves our well-being. In many ways, then, these four aspect of human nature clearly help to make us well-designed and successful creatures.

The problem is that these four (usually) good features of human nature are a mixed blessing. There are certain decision situations in which these same beneficial and even admirable aspects of human nature make for a tendency to fall into patterns of irrational choice. To the degree that we desire to be rational, and believe we ought to live up to standards of rationality, we must try to limit the influence of these parts of our nature on our effort to be rational. Being aware of these fallacies helps with this effort. They all have one common general form: an agent has a goal and the agent's menu of options contains at least one that, if chosen, would best achieve the goal compared to the other options in the menu, and the agent fails to choose it due to one or more of these features deep in human nature.

Due to the social nature of humans:

- Practical fallacy of decisional compromise
- Practical fallacy of decisional habit

Due to the experience of conflict and indecision:

- Practical fallacy of rash decision
- Practical fallacy decision passing
- Practical fallacy of decisional habit
- Practical fallacy of procrastination
- Practical fallacy random decision
- Practical fallacy of decision complication

Due to human weakness:

- Practical fallacy of sunk costs
- Practical fallacy of binding the will

Due to value distortion:

- a) The problem of relative value
 - Practical fallacy of relative value distortion
 - Practical fallacy of loss aversion
- b) The problem of discounting the future
 - Practical fallacy of unintended consequences
 - Practical fallacy of procrastination

Note that some of these fallacies appear under different categories. Even though they have the same name, they refer to different patterns of poor decision making. An agent falling back on his or her personal habits of making choices in order to escape conflict and indecision is not quite the same as an agent who chooses a course of action that has become his or her “normal” (that is: socially acceptable

and comfortable) way of behaving. Likewise, an agent who procrastinates because of indecision is not quite the same as an agent who procrastinates because of value distortion.

Also, even though we treated these practical reasoning fallacies one-by-one for learning purposes, we should be aware that agents are typically vulnerable to several of these practical reasoning fallacies while trying to make a decision. Imagine someone who faces a decision and procrastinates because of discounting the future and because of indecision. Now add into this mix the fact that such a person is probably subject to problems of binding the will and decisional compromise. You can appreciate that such an agent will find it difficult making a good decision (that is, without having learned some practical reasoning skills).

Being aware of these practical fallacies as patterns of irrational choices helps us recognize them in everyday concrete decision situations. This in turn can help us avoid making poor decisions ourselves as well as allow us to help others avoid them. But you might now be feeling that because these practical fallacies are so deeply connected to otherwise beneficial aspects of human nature, the effort to avoid them is bound to be a losing battle. Not so! The framework for making rational choices covered in this text can counter the influence of those aspects of human nature that make us prone to commit practical fallacies. The models, methods and principles of rational choice, if followed, make it difficult for such fallacies to disrupt practical reasoning.

There is a similar situation in the area of critical reasoning. If you are at all familiar with the rules of valid inference, it will be clear to you that they offer nearly foolproof protection from the common logical fallacies – provided, of course, that they are followed. Even though there are strong tendencies in human nature to commit the common fallacies, especially the informal fallacies, and even though these tendencies do not go away when we reason logically, it is impossible for them to get a hold on our reasoning if we carefully follow valid rules of inference. The same thing is true in the case of the methods of practical reasoning and principles of rational choice. If they are followed there is little room for practical fallacies to degrade our reasoning, in spite of the fact that all the features of human nature that make us

prone to these patterns of bad decision making remain with us as we reason. So the question is: to what degree will the principles and methods of making good choices be learned, practiced and followed?

EXERCISE:

1) Identify the practical reasoning fallacy in each of the following decision situations. (Take each situation as you read it, don't read more into it so that you start to see several possible fallacies being made.)

a) Sarah is offered a choice for her up-coming birthday between a puppy and a kitten as a pet. She likes them equally and can't make up her mind. As her birthday nears, Sarah asks her best friend Carol to make the choice for her.

b) New homeowners must decide on a color to paint the exterior of their house. They would love to paint it bright red, but instead decide on a neutral grey because they have moved into a conservative neighborhood and worry about neighbors disapproving of bright red. Whenever they see their house from a distance, they regret it isn't bright red.

c) Abby has put a lot of time and effort, as well as finances, into her new small business: a shop specializing in local hand-made jewelry, the kind that she herself loves. But things are not going well, and everyone with experience advises her to go out of business and cut her losses. Abby, however, can't bring herself to give up after all that she has put into her dream of owning a jewelry shop.

d) Paul has a term paper due Monday morning and realizes that he needs at least three days of solid work on it to get a passing grade, which he very much wants. It is Thursday and Paul has just

decided to go with friends on an over-night canoe trip all day Friday and Saturday. Paul wishes he had decided differently when he gets his paper back with a failing grade.

e) The only children of an elderly widow, a son and a daughter, must decide into which nursing-home they will place their mother. They are struggling with a difficult decision, and are caught in a cycle of gathering more-and-more information to help them make the right choice. They become so overwhelmed with all the details that they end up concluding that they are not able to make such a complicated decision.

f) Bill is running late to visit his father in the hospital recovering from surgery. He wants to bring flowers, but can't decide between roses or tulips, two of his father's favorite flowers. Annoyed at his indecision, he spontaneously buys a bunch of unknown flowers that he sees near the door to the flower shop.

g) Sally decides to drive an extra 5 miles to a gas station where gas is 10 cents cheaper per gallon. She fills up with 20 gallons and saves \$2.00. Next day, Sally is buying a new cell phone for \$199.00, and decides not to drive a half mile where she saw it advertised for \$197.00.

h) Sam signs up for a math course in his junior year, a basic freshman level course that he has been putting off but is required for a degree at his college. Sam believes that he is not good at math. Things start well in the course, but as soon as the material becomes the least bit challenging Sam decides to drop the course.

i) Jane is creeping along in heavy traffic on her way to work. There is construction ahead requiring all lanes to merge down to one. Jane decides not to let the car next to her cut in ahead of her, and maneuvers to close the gap. But the other driver thought Jane would let him in, and as he starts to merge their fenders hit. The damage was slight but costly, the other car was a high-end European model. 10 months later Jane's auto insurance is reviewing her policy, something Jane knows they do

every year, and raises her rates significantly based on her “fender-bender”. Jane now regrets her decision not to let that other driver merge in ahead of her.

j) Poor teeth tends to run in Pete’s family. He brushes once a day. One day he sees a sign in a drug store that says: “Brush after every meal, it’s the best way to keep your teeth.” He thinks for a moment and decides it’s too inconvenient to change his one-a-day brushing routine. A few days later Pete is in another drug store and reads a similar sign: “Don’t brush after every meal, it’s the best way to loose your teeth.” After reading the second sign, Pete decides on a daily oral hygiene change that includes brushing after every meal.

k) Sue has been invited to go dancing next Saturday night. She would love to go but is hesitant, because she thinks she can’t dance well and is afraid she’ll make a fool of herself. She puts off making a decision all week, even though she is asked two more time. As early Saturday evening approaches, Sue finally manages to overcome her fears and calls to say that she’ll be joining the group; but she learns that they already left for a bite to eat and will go directly to the dance from the restaurant.

l) Rob, recently promoted to upper management, is driving to the airport with two of his company’s chief executives; they are on their way to a business conference. Instead of the usual route to the airport, which Rob finds boring, he would like to try a new route. He considers the alternative routes briefly, and thinks to himself that if he were alone he would give one of them a try. Instead, Rob decides to stay with his company’s traditional route to the airport.

m) Beth feels trapped: this semester she is taking a history class that has tons of required reading, and a science class that requires a huge amount of study. The mid-term exam for each class is on the same day, and Beth feels that she only has time to study enough to do well on one of these, she’ll have to accept a poor mid-term exam grade on the other. But she can’t decide which one to neglect.

Rather than talk to her academic advisor, Beth makes her decision by picking a card from a shuffled deck: if red she'll work on her history class readings, and if black she'll study for her science class.

n) Smith's Hardware is a small family business that faces a major decision. Several large home improvement giants have built large stores in the area that have drawn away most of Smith's customers. In addition, the neighborhood has started to turn "seedy". The Smith family, however, can't decide what to do: go out of business by declaring bankruptcy, or sell the store to a group who intends to turn it into a sex shop. The family is equally uncomfortable with these options. Unable to choose the "lesser evil", the family decides to keep Smith's Hardware open because that what the family has "always done" during hard times.

2) Discussion topic: based on your own experience, which of the practical reasoning fallacies covered in this chapter might the following agents be especially vulnerable to. Why do you think so? Can you think of any other agents that might be particularly vulnerable to one or more of these fallacies?

- a) college students
- b) you yourself
- c) parents with young children
- d) government planners and policy advisors
- e) corporate executives
- f) medical professionals (doctors, nurses, EMT's, etc.)
- g) high school students
- h) elementary school teachers
- i) nations
- j) friends

Sources and suggested readings:

This chapter follows Mullen and Roth's (2002) lead in presenting several patterns of poor decision making in the context of psychological conflicts; see their Chapter 2 for a variety of interesting examples and for references to a range of psychological and applied economic literature on decision experiments. For value distortion fallacies, see Kahneman and Tversky's (1982) popular "Scientific American" article on framing effects, and especially loss aversion. Elster (1989a) offers a very accessible description of discounting in Chapter V, but Ainslie's (2001) influential work is highly recommended for anyone interested in the affect of discounting on our choices. For the problems of sunk costs and binding (constraint theory) Elster (1979) is considered a primary study, but his later *Ulysses Unbound* (2000) presents an updated treatment of a variety of binding problems and strategies. Part III of Holyoak and Morrison (2005) offers an excellent summary of the most important literature on a number of decision fallacies. Finally, for students with an interest in psychology, Plous (1993) provides a good introduction to a wide range of findings about decision making in various psychological and social situations. Practical reasoning fallacies form a large and growing area of study in the fields of behavioral economics, behavioral law, and evolutionary psychology; a web search of these fields will provide a variety of sources and links.