MAKING GOOD CHOICES: AN INTRODUCTION TO PRACTICAL REASONING

CHAPTER 11: PRACTICAL REASONING IN POTENTIALLY COOPERATIVE INTERDEPENDENT DECISIONS

In this chapter we continue examining decisions problems that are games. But instead of games in which agents are in a complete conflict of interest – the competitive or zero sum games of the last two chapters – we will now consider games in which agents can cooperate, at least to some degree, to achieve a goal. Here is an example.

1) Who calls back?

Two people are talking to each other on their mobile phones, one is walking and the other is driving. It is a heated conversation and it's turning increasingly unpleasant for both. Suddenly, they are briefly out of range of a transmission tower and are cut off. This was an important conversation; each wants to offer apologies and settle the disagreement. So they each wait for a minute, hoping that the other person will call back. But because both are waiting, each is not getting called back. Each starts to feel angry that the other person may not be willing to come to an agreement even though each feels personally willing to end the fight. After a couple of minutes, the person who was walking decides to take the initiative and call the other person back. However, just at that very moment the person driving likewise decides to take the initiative and call back. The result is that each can't reach the other, and comes to believe that the other is refusing to continue the conversation. Anger takes over. After a few more minutes, each starts to cool down and again tries to call back. But again it happens at the same moment resulting in failure. At this point, each person feels his good intentions are being rejected and decides in anger not to try the other person again but wait to be called back.

In this scenario, we see agents who do not have opposing goals; on the contrary, their goals overlap in such a way that both can have goal achievement. Also, we have two agents whose decisions are interrelated: how one agent decides depends on how the other agent decides and, as in zero sum games, this applies equally to each agents. In our mobile phone example, if the driver will be calling back, the walker won't; but if the driver won't be calling back, the walker will. And from the driver's viewpoint, she will decide to do the opposite of what the walker decides to do: call back if the walker doesn't, and not call back if she will. If only these agents can coordinate their decisions, if they can cooperate, both can have goal achievement. In the theory of rational choice, such interrelated – interdependent – decisions are not strictly competitive but potentially cooperative games. This is the topic we will look into in this chapter.

11.1 Potentially cooperative interdependent decisions

Potentially cooperative games are called "non-zero sum games" because the matrix contains at least one cell having outcome utilities that allow for a degree of goal achievement (or loss) for both agents; that is, one cell whose utilities sum to more than zero (or less than zero if the outcomes are negative). They are also called **potentially cooperative decisions** because in order to achieve the mutual gain outcome, the agents have to achieve some degree of cooperation or coordination with each other in their decisions. Thus, the agents are not necessarily opponents. We started this chapter with an example of two people trying to reestablish mobile phone contact. These two agents are not competing for a goal that only one can achieve; quite the contrary, each agent desires to achieve his goal in a way that allows, and perhaps helps, the other agent to achieve her goal. If the right cooperation is achieved in such decisions, both agents can "win". But if the agents fail to coordinate their decisions, each can easily lose part or all of their goals (the two mobile phone users don't reestablish contact). In cooperative games, then, there is potential for mutual goal loss as well as potential for mutual goal gain, and each agent has the power to cooperate with or to frustrate the other agent's decision.

Each agent, then, has basically two options: cooperate or defect ("defect" means not to cooperate or to choose the option that frustrates the other agent's goal achievement). If we use "C" to mean the cooperative option, and "D" to mean the defecting or non-cooperating option, the general matrix for 2-person potentially cooperative decisions is this:

		Col	
		C	D
Row	C	C, C	C, D
	D	D, C	D, D

In the theory of rational choice, these four outcomes have been characterized in a revealing way, and we will be referring to them in these ways as we work through examples. When both agents choose the cooperative option they gain (1) the **mutual cooperation** outcome (C, C), and when both agents choose the defection option both gain (2) the **mutual defection** outcome (D, D). But when one agent chooses to cooperate and the other agent defects, the cooperating agent gains (3) the **sucker's** outcome (for Row: C, D; for Col: D, C), and the defecting agent gains (4) the **free ride** outcome, also called the **temptation** outcome (for Row: D, C; for Col: C, D). As we shall see, given the goal sometimes an agent should choose to cooperate and sometimes should choose not to cooperate.

Five non-zero sum games are particularly interesting and in the theory of rational choice they have been, and continue to be, the focus of considerable attention. In fact, they have been given names. These five are thought to model important patterns of interdependent practical reasoning, patterns of cooperative decision making, and also model the danger of a breakdown in cooperation or outright obstacles to cooperation. They are in this respect sometimes described as "social dilemmas" and together are thought to get at the very heart of the problem of cooperative decision making. They will be the primary focus of attention in this and the next chapter. The names of these five games are: harmony, clash of wills, chicken, stag hunt, and prisoner's dilemma. This chapter will focus on harmony, clash of wills, and chicken; Chapter 12 will continue with the stag hunt and the prisoner's dilemma.   For each game:

1) We will first provide an idea of the pattern of cooperation that the game represents or models and we'll highlight the key question or problem in this pattern that we will be looking into.
2) Next, we'll look at one or more examples that serve to illustrate the game.
3) We'll then analyze the options and outcomes of each agent in the standard branching diagram.

This will mean assigning utility values to outcomes using the goal as a single criterion, and forming a rational preference order for each agent matching these utilities.

4) Last, we'll put each game into standard 2x2 matrix form to examine the interdependence of the decisions and to see if each agent has a rational choice. The methods of practical reasoning we will use are those that proved so successful solving competitive games: dominance, maximin reasoning (saddle point), and equilibrium outcome. (In non-zero sum games "equilibrium" is called "**Nash equilibrium**" after the mathematician John Nash who proved that every 2-person game contains at least one equilibrium outcome in either single or mixed strategies.) If it should turn out that these methods of practical reasoning fail to yield a rational choice, we'll consider what other additions to practical reasoning might gain the agents the goal.

These five games (harmony, clash of wills, chicken, stag hunt, and prisoner's dilemma) are special not only in offering valuable insight into forms of strategic practical reasoning for a decision to cooperate or not cooperate with others. There is another equally important, though troubling, aspect to some of these five games: as you will see, they test and expose the limits of practical reasoning. In contrast to the purely competitive zero sum games, all of which have rational choice solutions that clear methods of practical reasoning can discover, we will see that some of the cooperative games we'll be examining can't be solved by any of these standard methods of practical reasoning. These decision problems seem to have either no (single) rational choice solution (clash of wills and chicken), or if there is one by any of the standards methods that have served us thus far, it will look very unsatisfying and disappointing – even unacceptable (it will be sub-optimal) – to most reasonable people (stag hunt and prisoner's dilemma). And yet you will find these troubling cases to be important models of potentially cooperative human interaction. That practical reasoning should reach its limits and appear to give out just when it seems most needed to solve decision problems central to human cooperation is a valuable, even if discouraging, lesson to learn. If practical reason fails us in such important interdependent decisions, is there anything else that might help agents achieve the goal? Non-rational (but not irrational!) additions and alternatives to practical reasoning have been suggested and we will describe some of these for those cases where practical reasoning looks as if it has failed to reach and justify a rational choice. Let's now examine the first three of five games. We'll start with the least troubling one, harmony, and use it as our base-line to help us see how far the others depart from the satisfying rational choice solution it contains.

11. 2 Harmony

If two agents were friends and wanted to help each other, it is easy to see that they would choose to cooperate with each other for some mutual benefit. But what if they didn't know each other, or didn't like each other, perhaps are even enemies who distrust each other? What if each agent were choosing his goal purely out of self-interest, each wanting the goal only for himself? And let's suppose they are not forced to cooperate by a third party, say, an authority or power they both submit to such as a boss or a ruler. What form of practical reasoning would lead them willingly to cooperate and thereby help each other's goal achievement? Harmony represents such a decision situation. Any number of decision scenarios could be used to illustrate harmony. Here is one.

Suppose two neighbors, Robin and Cathy, are not getting along. Each would like to put a fence up on the boundary between their yards. They arrange to do this together for two reasons: each looks forward to having the fence completed (they have had it with each other!), and they distrust each other to position the fence exactly on the boundary line between their yards. The agreed Saturday morning to do the job arrives and they get under way, but things are going more slowly than each thought it would. Mid-afternoon comes and they are only 2/3 finished. They are getting tired, and besides that a tempting new goal enters the picture: it's time for a favorite TV sports program that each agent enjoys watching and rarely misses. What to do? They can both stay and finish the fence (mutual cooperation), or both quit to watch TV (mutual defection), or one can quit to watch TV (free ride) while the other agent stays and finishes the fence alone (sucker). Given each agent's goal of having the fence up, they want to keep working more than watching TV, but would rather watch TV than keep an eye on the sneaky neighbor if she'll finish the fence alone. What is the rational choice for each neighbor: stay on the job or quit to watch TV?

Agent	options	states	Robin's outcomes	ordinal rank	utility (1…10)
Row: Robin	stay	Cathy stays	finish fence together	4	10
		Cathy quits	Robin finishes fence, Cathy watches TV	2	3
	quit	Cathy stays	Cathy finishes fence, Robin watches TV	3	8
		Cathy quits	fence unfinished, both watch TV	1	1

Agent	options	states	Cathy's outcomes	ordinal rank	utility (1…10)
Col: Cathy	stay	Robin stays	finish fence together	4	10
		Robin quits	Cathy finishes fence, Robin watches TV	2	3
	quit	Robin stays	Robin finishes fence, Cathy watches TV	3	8
		Robin quits	fence unfinished, both watch TV	1	1

Given the goal, it is clear that each agent desires mutual cooperation to a free ride, a free ride to being a sucker, and being a sucker to mutual defection. This preference order defines the game **harmony**. Now we'll put this particular example of harmony into matrix form to see what choice practical reasoning can justify.

Goal: both agents have the same goal – a fence between their yards

		Cathy: C1: cooperate (stay)	Cathy: C2: defect (quit)
Robin:	R1: cooperate (stay)	4, 4 (10, 10)	2, 3 (3, 8)
	R2: defect (quit)	3, 2 (8, 3)	1, 1 (1, 1)

Each outcome cell includes both the outcome utilities (assigned to outcomes using the goal as a single criterion), and the ordinal rank of the outcomes (directly above outcome utilities). It will sometimes be convenient to use just the ordinal rank numbers to analyze and evaluate cooperative games. Note that every cell offers some opportunity for mutual gain; that is, no cell sums to zero. Yet the matrix shows clearly that practical reasoning, in this case in the service of pure self-interest, moves the agents to the mutual cooperation outcome (10, 10) and that this is a <u>Nash equilibrium</u> from which neither agent will switch until each achieves the goal. For Robin, R1 <u>dominates</u> R2; as Robin ponders quitting rather than staying on the job, she realizes that she will do better in her goal achievement with R1 no matter what Cathy does. The "pull" toward R2 is not strong enough to overcome the R1 outcomes. So, as much as she might enjoy watching her favorite TV sports program, Robin – if she makes the rational choice – will continue putting up the fence even if Cathy quits. Likewise for Cathy: C1 dominates C2. Mutual cooperation results and each agent ends up (unintentionally!) "helping" the other gain the goal. Note, however, that maximin reasoning does not solve this decision problem; there is no saddle point outcome pair.

It is important to see that harmony is not a social choice situation. Robin and Cathy, in the above story, are not two members of a group each contributing a decision to a collective or group decision whose outcome will benefit its members. Each is making an individual decision; neither intends to help the other achieve her goal; neither has the other as a stakeholder in the decision to cooperate. However, their goals happen to coincide in a way that goal achievement on one agent's part necessarily and automatically promotes goal achievement for the other. They have an <u>interest in common</u> (each desires to isolate herself from the other by way of a fence separating their yards), but do not have a <u>common interest</u>, as this was defined in Chapter 1. This important feature of harmony can perhaps be brought out even more clearly than the story of Robin and Cathy illustrates if we alter the story slightly and imagine that they are perfect strangers who know nothing about each other and who work on the fence not together but, let's suppose, on alternate days. The goal remains the same: to separate their yards with a fence; only now think of these two agents as having never met and as having no personal interest in each other at all. As agents "playing" harmony, they still cooperate if each chooses rationally, rather than quit each mid-afternoon to watch TV.

The larger lesson we can take away from harmony is that sometimes (but not always) people acting out of nothing but rational self-interest will achieve mutual cooperation and thereby help each other maximize utility; their decisions are “locked” together in this special pattern called harmony. In such situations they don’t need any “internal” additions to practical reasoning to help them discover cooperation as the rational choice – additions such as a feeling of altruism or an emotional attachment to each other to explain why they are “getting along.” Likewise, they don’t need an “external” third party power – a boss or a ruler – to keep them cooperating and to make sure they don’t interfere with each other, or exploit each other, or frustrate each other’s goal achievement. This seems like an interesting and an important lesson contained in harmony.

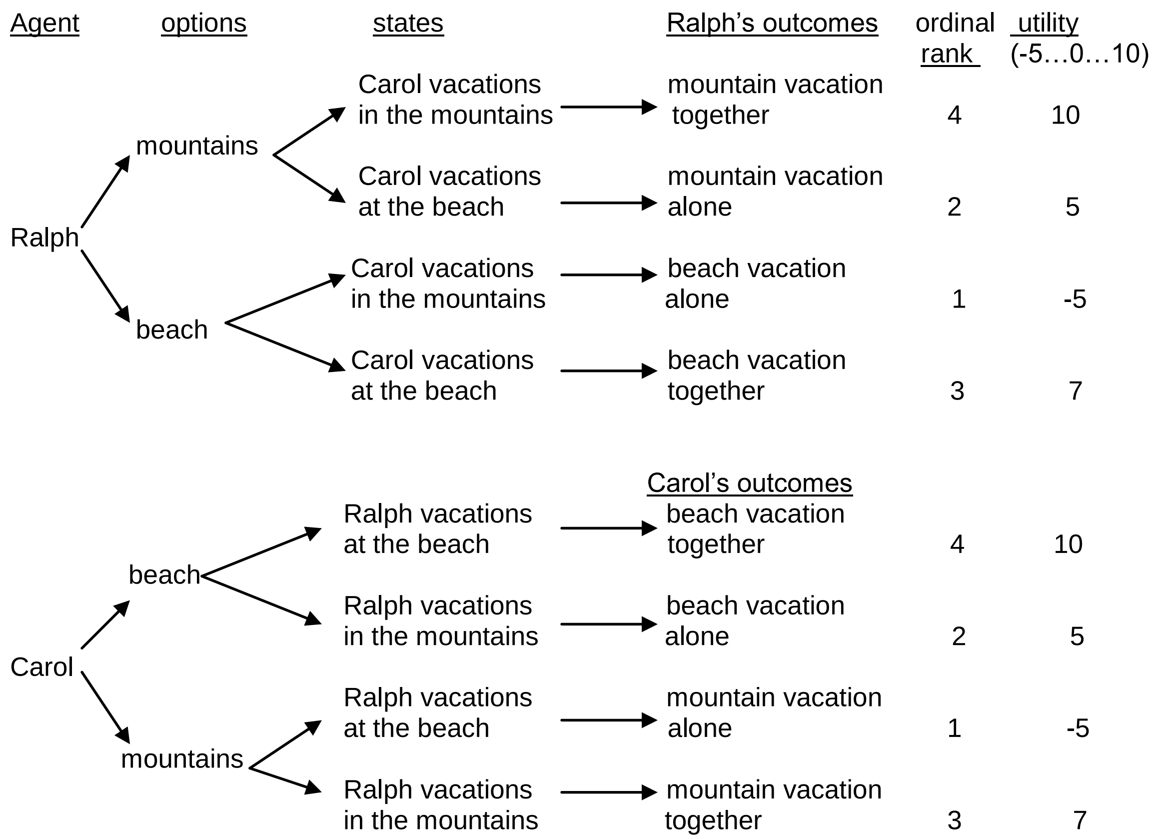
11.3 Clash of wills

Is mutual cooperation ever irrational? Is it ever rational to defect, to stand one’s ground and refuse to cooperate, and gain a “free ride” by taking advantage of another agent’s willingness to cooperate? The game known as “clash of wills” provides a model that helps us investigate these possibilities. Think back to the decision problem that opened this chapter: two agents trying to reestablish mobile phone contact. In order to reestablish contact both can’t cooperate (call back) and both can’t defect (not call back), one has to cooperate and one has to defect. Only this pattern of decision coordination and not the other two gains each agent their goal. The questions, however, is: who calls back and who doesn’t? If both do (at the same time) or if neither do, then there is no goal achievement.

Here is a classic clash of wills. Two people very much in love desire to marry, but they differ in religious belief: one, let suppose, is very religious but the other is an atheist. The believer naturally wants a religious wedding and considers a civil wedding not a true marriage. The atheist wants a civil wedding, and considers a religious wedding a waste of time and money. In order to have goal achievement (marriage) one agent, and only one, must give in and do things the other agent’s way.

As these examples illustrate, the game “clash of wills” models coordination problems, decision problems that easily lead to total goal loss if coordination fails. The following story is another classic clash of wills. It will serve as our decision example to analyze and evaluate, but any number of scenarios could serve just as well.

Ralph and Carol are planning a vacation together. Each strongly desires to be with the other more than to vacation without the other, but Ralph wants very much to go to the mountains and doesn't like the beach while Carol would love to vacation at the beach and hates the mountains. They discuss their options; each tries to persuade the other to vacation at his or her preferred place but each is getting nowhere with the other. The discussion gets heated; each is becoming more-and-more stubborn and unwilling to give in. In anger they stop speaking to each other and yet they must decide on a vacation spot. What is the rational choice for each agent?



It is clear from this analysis that each agent desires a free ride to being a sucker (giving in), being a sucker to mutual defection, and mutual defection to mutual cooperation. This preference order defines the game **clash of wills**. Now let's put this particular example of clash of wills into matrix form and see if practical reasoning can justify a rational choice.

		Col: Carol	
		C1: cooperate (mountains)	C2: defect (beach)
Row: Ralph	R1: cooperate (beach)	1, 1 (-5, -5)	3, 4 (7, 10)
	R2: defect (mountains)	4, 3 (10, 7)	2, 2 (5, 5)

No agent has a dominant option. Can maximin reasoning lead to a rational choice? There is a saddle point cell, it is mutual defection (5, 5). But note that it is unstable. It is not a Nash equilibrium cell; each can do better (more goal achievement) by switching to the cooperative option, providing that the other agent refuses to switch to the cooperative option and remains defecting. Both should not choose to cooperate, for the mutual cooperation outcome is the last thing these agents desire given their goals! It would be irrational for each to cooperate, and also irrational for each to defect, though not as irrational. The Nash equilibrium cells, from which neither agent has any practical reason to switch, are the two that sum to maximum utility resulting from the choices (R2,C1) or (R1,C2). The rational choices, then, is for one agent to hold out for their most preferred outcome (defect) and the other agent to give in and settle for their second best outcome (cooperate). The problem, however, is that there are two equilibrium cells yielding equally maximum joint utility, one favoring Ralph (R2,C1) and the other favoring Carol (R1,C2). One agent must give in and cooperate, but the other can't give in and must insist on having things "my way," if one of these best outcomes is to be reached. This, then, is the clash: who should compromise? Who should give in? Who should not give in? One must compromise and the other must refuse to cooperate or both lose their goal.

So far, it looks as if a clash of wills has no single rational choice solution. Practical reasoning finds two equal solutions by equilibrium reasoning and no other method of practical reasoning breaks the deadlock. You might disagree and think that communication is the answer. If only agents caught in a clash of wills talk things out, you might think, it would bring about the needed compromise on one of their parts. But, as the story of Ralph and Carol is meant to show, communication is very much part of the clash, not necessarily the way to a solution. If increased attempts to pressure the other agent to "give in" and do things "my way" causes increased resistance and equally strong attempts at counter-pressure, communication could easily increase levels of stubbornness and anger and end up causing each agent to alter their goals. To overcome perceived stubbornness, suppose each agent resorts to threats (a form of communication!) to "go it alone" or resorts to issuing ultimatums. Suppose, for example, Carol storms out and her last words are "Ralph, I'm going to vacation at the beach, you can join me or not – it's up to you!" Instead of a vacation together, communication could escalate the clash into a desire to vacation alone – in effect ending the relationship.

In this case the outcomes, with only an ordinal rank, would be:

Ordinal rank	Ralph's outcomes	Carol's outcomes
4	mountains alone	beach alone
3	beach alone	mountains alone
2	mountains with Carol	beach with Ralph
1	beach with Carol	mountains with Ralph

The decision matrix is clearly no longer clash of wills. Let's call it "severed relations."

		Col: Carol	
		C1: mountains	C2: beach
Row: Ralph	R1: mountains	2, 1	4, 4
	R2: beach	3, 3	1, 2

There are two Nash equilibria in the game severed relations: outcome (3,3) resulting from the choices (R2,C1), and (4, 4) resulting from the choices (R1,C2). But given the new goals of these former friends who now no longer desire each other's company, it is clear that (R1,C2) is the rational choice; it is the best outcome, while (3,3) is sub-optimal. The danger of continued communication between agents in a clash of wills, then, is that the more opportunity each defecting agent has to asks (pressure?, demand?, threaten?) the other to cooperate (give in and become the sucker), the less each agent appears to the other worth cooperating with. Communication can easily transform a clash of wills into severed relations; it is better for the agents not to communicate if it looks as if this is the direction communication is headed. In the story, Ralph and Carol wisely <u>stop</u> talking to each other.

You might be asking: If communication is not always the answer, if it contains at least as much danger as hope, then isn't overcoming a clash of wills simply a matter of priorities? Doesn't Ralph have to ask himself, which is more important: Carol or the mountains? And doesn't Carol have to ask herself, which she values more: Ralph or the beach? Once each realizes (let's suppose) that the other person is more important than a vacation spot, isn't the problem solved? Clearly not! To achieve the goal, it can't be the case that each gives in to the other. One of these agents has to say that his (her) vacation spot is more important than the other person (and not compromise), and the other agent must say that the other person is more important than his (her) vacation spot (make the compromise). Getting one's priorities straight, then, will not solve the problem, if each values the other person more than the vacation spot; it just shifts it into different terms.

11.3.1 One clear rational choice solution to the clash of wills would look like this.

		Col: Carol C1: cooperate (mountains)	C2: defect (beach)
Row: Ralph	R1: cooperate (beach)	1, 1 (-5, -5)	3, 4 (7, 13)
	R2: defect (mountains)	4, 3 (10, 7)	2, 2 (5, 5)

This matrix says that <u>both</u> agents accept that the beach vacation together means more to Carol than the mountain vacation together means to Ralph (perhaps because going to the beach is desired more by Carol than going to the mountains is desired by Ralph). The two Nash equilibrium cells are no longer equal, the (R1, C2) outcomes sum to more total goal achievement than the (R2, C1) outcomes. In other words, it is a bigger loss of goal achievement – a bigger disappointment – for Carol to cooperate and Ralph defect (she goes from 13 to 7) than it is for Ralph to cooperate and Carol to defect (he goes from 10 to 7). There is a mutually acknowledged lack of symmetry, an asymmetry, in the outcome utilities. In this situation, it is rational for Ralph to give in and for Carol to stand her ground; the rational choice is (R1, C2). When this kind of solution is available, the game is a weak clash of wills. In a **weak clash of wills**, then, a rational solution depends on a special form of common knowledge: a common mutual acceptance. It is accepted by both agents (and accepted by each that it is accepted by both, …) that the value of goal achievement (represented by the maximum utilities) of each agent are unequal; the agent whose goal achievement is valued less cooperates and the agent whose goal achievement is valued more defects.

Suppose there is no such common acceptance? Suppose that each agent claims that his best outcome is worth just as much (or, in an effort to make the other agent give in, worth more) to him as the other agent's best outcome is worth to the other agent? This situation is a strong clash of wills. A **strong clash of wills**, then, allows no rational choice based on a common mutual acceptance of unequal goal values. Let's divide the strong clash of wills into two categories: those that happen between agents only once (the **one-time strong clash of wills**), and those that happen repeatedly between the same two agents (the **iterated strong clash of wills**). Let's look at the iterated case first.

There are two ways that the iterated strong clash of wills is thought to have a rational solution: (1) by agreement and (2) by convention. Both are based on the principle of equally distributing the cooperate-defect options between the agents. The idea is that, as the clash of wills is repeated again and again, if each agent gives in to the other an equal number of times, neither agent's willingness to cooperate is being taken advantage of by the other agent.

(1) In the method of agreement, the agents agree to alternate choosing the cooperation (or the defection) option. For example, if Ralph and Carol will be vacationing together many times, it is rational for them to take turns between mountains and beach. However, there are two problems with solution by agreement that practical reasoning must overcome. One problem is how to start the iterations: will the first time be mountains or beach? The agents could leave this decision up to chance, say, by flipping a coin. Or, because the cooperating agent will be the defector next time around, it becomes easier to give in and so one agent could volunteer to be the first to give up his/her first choice in the spirit of compromise and peace-keeping. But note that flipping a coin or peace-making would be recourse to a non-rational method of arriving at a solution, required because practical reasoning can go no further discovering how the iterations should start. If the first round can't get started, and it can't by using methods of practical reasoning only, then clearly there is no (initial) rational solution by agreement to the iterated clash of wills.

The second problem with agreement is more troubling. The agent who cooperates on the initial clash of wills must, of course, trust that there will be a second round of the same clash of wills. But this agent should not also have to (merely) trust that in the second round the other agent who now defects will keep to the agreement to cooperate. In other words, practical reasoning requires not just an agreement, but a **binding agreement**. What's to keep the agent who has things go "my way" the first time from changing his mind, reneging on the agreement, when it is time to cooperate in the second round of the clash of wills? Worse, how can the agent who is willing to cooperate in the first round be sure she has not been duped by an empty promise that next time around she will get her free ride and the other agent will become the sucker? You may think that Ralph and Carol in our story are honest agents who desire each other's company. Keep in mind, however, that in an iterated strong clash of wills individual agents Row and Col might be two corporations headed by CEOs with very big egos, or two nations headed by ruthless leaders each of whom sincerely believes that his or her country is second to none, or two people who don't have sweet personalities and instead are very much "out for themselves" in dealing with each other.

What might bind an agent to an agreement to cooperate in the second round, once this agent has defected in the first round? Ultimately there seems to be just two forces that make an agreement binding: (i) fear of consequences (for example, fear of penalties from a third party “enforcer” if the agreement is formal, or fear of severed relations and damage to one’s reputation if the agreement is informal) and (ii) the force of morality (for example, the moral duty to keep one’s word, or the moral virtue of honesty). Whatever the binding force, the ideal rational agent will require a binding agreement in an iterated strong clash of wills (and – as we well-know – real agents must often rely on trust and realize that they are vulnerable to getting suckered).

But now note what has happened. Any agreement that is sufficiently binding on agents in an iterated clash of wills introduces strongly feared or morally unacceptable negative consequences (that is, disutilities) into the game for breaking the agreement. This, in effect, changes the game from the iterated strong clash of wills into another one having a different mix of outcome utilities, a game that has these new disutilities entered into the outcome matrix. A binding agreement, then, does not so much <u>solve</u> the iterated clash of wills as <u>change</u> it into another game. This means, in effect, that the original iterated strong clash of wills can’t be rationally solved by agreement. Either the agreement is not binding in the original game (and so does not amount to a rational solution), or it is binding and the game is no longer the iterated strong clash of wills (and again this is not a rational solution of <u>that</u> game). It seems that here practical reasoning has reached a limit; agents in this game cannot depend on practical reasoning to justify a rational choice solution, but must turn to “outside” help in the form of some kind of reliable, enforceable – i.e., binding – agreement.

(2) Now let’s consider a rational solution by convention. Iterated strong clashes of wills represent, as we have seen, coordination problems. Goal achievement requires that the agents work out some form of equal distribution in choosing the cooperation-defect options among themselves. Instead of alternating the options between Row and Col, as a solution by binding agreement requires, suppose we fix the options by an unwritten rule (let’s say, arbitrarily, Row always cooperates and Col always defects) but now we let the agents switch back and forth between the

Row “position” and the Col “position”. Let’s say that everyone who ever takes on the Row “role” in a clash of wills conforms her behavior to cooperation as if the decision were already made, and likewise everyone who is in the Col “role” conforms to the defect option behavior. Everyone expects this of everyone else, it is common knowledge, and all the individuals involved get to be Row agents and Col agents roughly an equal number of times. Such a “self-sustaining” situation is a **convention**. Here are some examples.

(1) Two people approaching each other in a hallway don’t collide because each conforms to the convention: stay to the right. But when the hallway is very narrow, not wide enough for two average size people to pass each other, the convention might be: the first to enter the hallway stays in the middle and the second to enter moves out of the way. The idea is that for all those who use the hallway, any given agent will enter before and enter after another agent roughly an equal number of times, and so the convention distributes free ride and sucker’ outcomes to agents in roughly equal amounts.

(2) Consider the two people we imagined at the beginning of this chapter who desire to reestablish mobile phone contact. The convention might be: the person who called first calls back. Again, the idea is that a mobile phone user will over time be disconnected roughly an equal number of times as a caller and as the one called, so cooperation and defection get distributed among mobile phone users more-or-less equally by such a convention.

(3) Think of the convention at a 4-way intersection with stop signs. To go first while the other drivers wait for you gives you the free ride, and they the sucker’s payoff. Clearly not all drivers approaching such an intersection can cooperate (what would happen?), nor can they all defect (why not?). In the United States the convention that solves this clash of wills is: the driver on the right goes first (defects) and the one on the left waits (cooperates). This is “fair” because you will approach such intersections in your life as a driver more or less equally as the one on the left (Row) and as the one on the right (Col); that is to say, you will get about as many free rides as

you will be the sucker, and likewise for all the drivers who use the 4-way intersection on an ongoing basis.

(4) Here is another driving example. Have you ever been in a roadway construction zone or a traffic accident area where two and sometimes three lanes must merge down to one? Let's say that you must merge into the left lane from the right lane. Who goes first, the other car or does the other car let you in ahead of him? One of you must let the other go first. In many areas of the country there is a well-known convention that solves this coordination problem: alternate merge, like a zipper. Again, the idea is that this convention will distribute the free ride and the sucker's outcomes to each driver in roughly equal amounts over time. But in other parts of the country the alternate merge convention never caught on. In such areas, you must either wait to merge until a considerate driver lets you in or sometimes you must force your way in by threat of collision, making the other driver cooperate or risk damage to her new car.

(5) As a final example, let's make up an iterated strong clash of wills and let's try to imagine a convention that solves it. We will stay with driving problems. Suppose in the future fuel prices get so high that more and more people having the same destinations seriously start to carpool. Agents desire to carpool, each owns a car, but let's suppose that no one wants to use his own car and do the driving; all carpoolers desire to be passengers (free-riders) equal in strength to their desire to carpool (their reasons might be things like keeping mileage low, saving on fuel bills, keeping repair costs down, etc.). The driver, then, is the agent who cooperates, and the passenger is the defector. This is a classic clash of wills. If everyone cooperates, everyone drive her own car and there is no carpooling. If everyone defects, no one drives and there is no carpooling. Some carpoolers must give in and drives their own cars, and others must refuse to drive and enjoys the free ride. It is a strong clash of wills because, let's imagine, no agents mutually accept that being a passenger is more valuable to any special person or group than it is to any other person or group within the population of carpoolers; everyone equally desires not to be the one who carpools driving his own car. It is an iterated clash of wills because, let's imagine,

there are a variety of destinations that different people desire to carpool to on an ongoing bases. What convention might naturally emerge from all these imagined carpool interactions that rationally solves this clash of wills? To simplify this decision problem still further, imagine that people carpool only two at a time, but not the same two each carpool trip. What are the pros and cons of this convention: the person with the newer car drives? What's wrong with: the younger person drives? How about: the person with the most fuel-efficient car drives? Finally, what about: the person who was a passenger last time now drives? Can you suggest a better convention?

Let's clearly note what a convention accomplishes. Consider the above examples carefully. Following a convention in effect <u>removes</u> the opportunity to make a decision; there is no need for agents to discover a rational choice because the convention already "pre-decides" the matter. The convention assigns the Row option or the Col option to every agent in the decision situation. If the convention is, say, alternate merge, then neither you nor the other driver have to worry about who "gives in" and who insists on having it "my way." As in the case of solution by binding agreement, this is <u>not</u> a rational solution to an iterated clash of wills by a method of practical reasoning; it is instead a way of avoiding the decision problem altogether by taking it out of the agents' hands. Here again, practical reasoning seems to have reached its limits and agents must turn to the "outside" help of a convention for a solution.

11.3.2 The hardest clash of wills case is the one-time strong clash of wills. It is rational to avoid such decisions, if possible. The weak clash of wills (whether one-time or iterated) has a rational solution based (by definition) on a mutually accepted asymmetry or inequality between the agents. The iterated strong clash of wills ideally can be avoided by a convention or avoided by a binding agreement (even though in reality such agreements and conventions might be hard to achieve). None of these mechanisms, however, apply to the one-time strong clash of wills, in which the danger of severed relations is strongest. If Ralph and Carol must decide only once on a vacation spot and can't take turns, and each is equally strongly committed to a different favorite place (mountains or beach), and neither will back down, then they have reached an impasse that practical reason can't resolve. They might agree to let chance decide the matter and flip a coin, if communication between them has not completely broken down. One might give in out of fatigue or out of a moment's weakness as the stressful attempt to pressure each other gets carried out. One might somehow trick the other to accept the sucker's outcome. But it is important to see that none of these possibilities is a <u>rational choice</u> that results from applying methods and principles of practical reasoning. These are non-rational ways to supplement practical reasoning so that some kind of solution is arrived at even though it is not a rational choice solution.

11.3.3 Before turning to the next potentially cooperative game, let's note an important general features about the clash of wills. Contrasting this game with the previous one, harmony, will help to highlight this features. In harmony, 2 agents have an interest in common, the goal. When we analyze the goal, we see that it is simple; it contains just one objective. What creates the decision problem is that before the goal is achieved, another simple goal enters the picture and competes with it. This second goal is logically incompatible with the first goal; they are mutually exclusive. In our example of harmony, two neighbors desire to fence themselves off from each other (the primary simple goal) but before this goal is achieved, a second simple goal tempts each agent to defect (a relaxing TV sports program). Each agent can achieve the fence or the TV program, but not both at the same time. However, the desire of each agent for the second goal is not strong enough to cause mutual defection. Thus, the agents in harmony rationally choose to cooperate.

The goal-structure of the clash of wills is quite different. Each agent has a complex goal having 2 objectives. For one objective, the agents have a <u>common interest</u> (not just an interest in common), making the clash of wills in this respect a social choice. In our example, both Ralph and Carol have a common objective to vacation together. In this they are each a member of a society (of two), and the decision of each contributes to a joint or collective decision whose objective benefits the "group." With respect to this objective, the agents are in harmony. The other objective, for Ralph it is a mountain vacation and for Carol it is a beach vacation, is for each completely compatible with the first objective, but incompatible with each other. This is what creates the decision problem. They can't vacation at the beach and at the mountains at the same time together. It is with respect to the second objective that the clash of wills is an individual decision for each agent and not a social choice. A brief goal analysis will help make this feature of the clash of wills clear.

Goal (Ralph)	Objectives (1...5)	Attributes (1...5)	
A good vacation (1.0)	vacation together	go to the same vacation spot	.5
	vacation at a pleasant place	go to the mountains	.5
			1.0

If we do the same goal analysis for Carol, everything is the same except that the second attribute is beach instead of mountains. Adding this attribute will give a sum of 1.5 (.5 for the same vacation spot which is their common interest, plus .5 for Ralph's mountains, plus .5 for Carol's beach = 1.5). This clearly violates the rule of goal analysis and option evaluations that requires criteria to sum to the same value as the goal, namely 1.0. There is .5 excess value that must be eliminated in order to have goal achievement. Whose .5 gets eliminated, Ralph's or Carol's?

11.4 Chicken

Sometimes it is hard to ignore or walk away from a challenge. Suppose the school bully picks on a victim in front of peers; backing down might be worse than taking the bully's abuse. Or, suppose someone challenges your truthfulness on an issue in front of your friends; defending yourself might be better than ignoring it in the hopes of avoiding a confrontation. Standing up to a challenge, showing your willingness to tough it out, not being a quitter, and not being the first to back down or show weakness or be intimidated when others are looking to judge your worth: these are qualities that are widely admired. When a person shows the opposite qualities, weakness and a willingness to be the first to give up when confronted with a challenge, it is commonly looked down upon. Also, in certain social interactions giving the appearance of strength is sometimes a good way to save yourself later trouble from anyone who might think you can be taken advantage of, or be easily intimidated, or have a readiness to back down (in the same way that appearing weak and vulnerable is sometimes a good way to gain assistance and sympathy). Chicken is an interdependent decision problem that helps us look at these two strategies: standing up to a challenge verses giving up. There is an interesting aspect about meeting a challenge that chicken also serves to highlight, namely the role of deception, bluffing, and intimidation as non-rational additions to the decision problem when practical reasoning is not able to arrive at a rational choice solution. Let's first consider deception.

People deceive one another all the time. I don't mean outright lying, although that too. I'm referring to situations in which people try to appear to each other as <u>other</u> than what they really are. Sometimes people pretend to be the very opposite of what they are. This is typically done by behavior patterns as well as with words. So, for example, a person who is telling a lie tries to appear honest to his victim. Or, a person who is afraid of something (let's say she is giving a public speech) tries to appear to the audience unafraid and confident. The opposite might happen: a skilled public speaker might think that the best way to win an audience's trust is to appear nervous and so might start his speech in a hesitating way saying, "I'm really nervous so please forgive my fumbling.", and by this "moment of honesty" could trick the audience to be more accepting of his position than it would otherwise be. Or, someone who is ignorant about an issue tries to appear knowledgeable, or perhaps the other way around: a person who knows certain important things might find it best to play dumb. Or, someone who is full of anxiety and self-doubt, say during a job interview, tries to appear calm and self-confident. A person who is feeling sad or in pain might try to hide it with a smile, or an angry person might try to cover up her feelings by putting on a façade of being unconcerned. In many sports events and contests it is expected that the weaker person or team will try to appear to their opponent to be the stronger in an effort to intimidate and "soften up" the competition. These are all cases of deception, and as you can see many might be perfectly acceptable given the circumstances, as for instance in the case of the job interview or the sports event. Experience tells us, then, that the art of deception is widely practiced; duplicity, subterfuge, camouflage, pretence, and exaggeration are common strategies people use in trying to get what they are after.

The questions for us are: in making rational choices, is it ever a help to one's rationality to appear irrational? Is it ever strength to appear weak and to appear as if one's hands are tied? Chicken is a potentially cooperative game that let's us examine this interesting form of deception, and let's us do so in a particular context: being challenged. Before analyzing this game in its basic form, here are some examples to consider.

Imagine two college students at a party who get into a dispute about who can drink more alcohol. They decide to settle the issue with a beer-drinking contest to see who quits first. All their friends are watching and cheering them on; in fact their duel has become the center of the party. They drink to the point where each feels he can't take any more beer and wants to quit, but each realizes that quitting means losing the contest, and looking weak and foolish in front of everyone. As they continue drinking beer, each is hoping the other will quit at any moment, and each pretends he is not about to quit and even brags that he can handle a lot more beer.

Imagine two automobile companies that are competing for car sales by lowering prices. They are in a price war. If they keep lowering car prices, however, both companies will suffer (they will have to lay-off workers and could even go bankrupt); but the one that stops first to lower prices will suddenly lose a lot of business to the other that lowers a bit more, because there will be a big swing in car sales away from the higher priced cars and toward the lower priced ones. Each company tries to give the appearance to the other that it can afford to lower prices a lot more.

Imagine two nations with a long history of war between them. Because of the threat each nation feels from the other, they use more and more of their national budgets to build defense forces: armies, planes, missiles, bombs in huge numbers, etc. They are in an arms race. If this military buildup continues, however, it will devastate the national resources of each nation; but the first nation to halt its military arms race will be at the mercy of the superior military forces of the other nation. Each publicly announces it is willing and able to support another round of massive military buildup, even though in reality each is already stretched to the limit and have populations starting to suffer.

Picture two people at a local fair who have entered a pie-eating contest. The race is on and each has reached the point where eating more pie will literally make them sick, and yet the first one to stop will lose the prize of $100 and disappoint their children and friends who are cheering them on. Each tries to make the other think she is enjoying the contest enormously and can eat a lot more pie.

As a last example, imagine two men pursuing the same woman. Each is trying to impress her and win her over by taking her out to expensive clubs, buying her expensive gifts, and being very lavish. Neither can afford to continue these expenses, yet each feels that if he lets up and fails to outdo the other in lavish dates, he will be rejected in favor or his rival. Each tries to give the impression to his rival that he can and will continue being lavish.

What do these 5 examples of chicken have in common? (1) They are all interdependent decision problems in which each agent faces a challenge and has two options: either quit or remain in a potentially harmful course of action. (2) The agents are on a “collision course” with some negative consequence each desires to avoid; so, each wants to quit but wants the other to quit first.
(3) Deception, bluff, and intimidation play an important role in trying to force the other to quit first.
(4) The goal goes to the last one who quits, and the accusation of weakness (chicken!) goes to the agent first to quit.

Chicken gets its name from a dangerous game that teenage drivers are said to sometimes play, and it is standard to illustrate this decision problem with this teenage game. Here is one possible way the challenge goes.

Two teens, trying to impress their friends and gain acceptance, brag about their ability to face danger, and each calls the other a chicken. They challenge each other to a driving duel to see who will chicken-out first. Each accepts, for neither thinks he will be the one to back down and lose face. On a straight empty road these two teens will speed neck-and-neck toward a large stationary cement structure (or in some versions, speed toward a cliff). The contest is to go faster and faster, aiming the cars at the structure (cliff) to see who will break to a stop or swerve first. The one who stops or swerves first is the chicken. Many peers are watching, ready to cheer the fearlessness of the winner and ridicule the chicken.

In analyzing chicken, we give each agent the same general goal: to demonstrate superiority (in whatever the specific challenge or contest or rivalry might be). To cooperate is to quit the course of action first and give up the goal of superiority in favor of another goal: safety, or avoiding harm (in the case of the teenage car duel, stopping first avoids potential death). To defect is to continue on the course of action trying to gain the goal: acknowledged strength, or superiority (in the teenage car example, driving straight ahead toward the cement structure or cliff and stopping only after the other stops).

Agent	options	states	Row's outcomes	ordinal rank	utility (-10…0…10)
Row	stay the course	Col stays the course	harm (perhaps death)	1	-10
		Col stops first	winner (superior)	4	10
	stops first	Col stays the course	chicken (loser)	2	- 9
		Col stops first	tie (neither is chicken, neither is superior)	3	5

Agent	options	states	Col's outcomes	ordinal rank	utility (-10…0…10)
Col	stay the course	Row stays the course	harm (perhaps death)	1	-10
		Row stops first	winner (superior)	4	10
	stops first	Row stays the course	chicken (loser)	2	- 9
		Row stops first	tie (neither is chicken, neither is superior)	3	5

We can see from this analysis that each agent desires a free ride to mutual cooperation, mutual cooperation to being a sucker, and being a sucker to mutual defection. This preference order defines the game **chicken**. Now let's put this particular example of chicken into matrix form and see if practical reasoning can justify a rational choice.

		Col: C1: cooperate (stop first)	Col: C2: defect (stop last)
Row	R1: cooperate (stop first)	3, 3 (5, 5)	2, 4 (-9, 10)
Row	R2: defect (stop last)	4, 2 (10, -9)	1, 1 (-10, -10)

You can see that neither agent has a dominant option. Also, maximin reasoning does not lead to a rational choice; there is no saddle point outcome cell. As with the clash of wills, there are two Nash equilibrium points in chicken: the outcomes resulting from (R1,C2) and (R2,C1). If one agent believes that the other will be the first to stop, it is much better to stay the course and stop last (and gain utility 10: the goal). And if one agent believes the other will stay the course to the bitter end, it is better to stop first (and suffer -9 disutility rather than -10 disutility). Note that being a sucker is bad, but not quite as bad as what could happen if there is mutual defection. In our story, the teenage drivers must weigh the fear of crashing into the cement wall (serious injury and perhaps death, worth -10) against the fear of being mocked by peers as a chicken (worth -9). Neither is a very appealing outcome, but one is worse (more undesirable).

Chicken might at first seem like a zero sum game in so far as goal achievement by one agent means goal loss for the other. But the utility numbers show that this is not a zero sum game. While it is true that each agent can't quite "have his cake and eat it to," as is sometimes the case in the clash of wills, nevertheless both can end up achieving mutual cooperation (the agents tie, are chicken to the same degree, and so each avoids looking inferior to the other), or equal goal loss in the case of mutual defection. If a prize is at stake, as in the above example of the watermelon-eating contest, a tie would mean that neither gets it (or perhaps the agents can split the prize). Even the free-ride and sucker's payoff don't quite sum to zero, though it's close, at least in the way the game has been analyzed here.

Why, then, isn't mutual cooperation the rational choice? <u>Given the goal</u>, and given the belief that the other agent will cooperate, it is clear what a rational agent should do: defect. Mutual cooperation is not a Nash equilibrium outcome, it is unstable; each can do better switching to the other option, providing the other remains cooperating. But if both switch, both end up with the worse outcome. Suppose, for example, the two teenage drivers secretly agree that each will stop at the same designated spot before hitting the cement wall (or going off the cliff) so that both could come out looking equally chicken. Should each trust the other to keep such an agreement? If Row really trusts Col to keep his word and stop at the agreed spot, then all the more Row has strong practical reason to defect and continue speeding beyond that spot (as always, given the goal; if promise-keeping is more valued than the goal, the game isn't chicken any longer!)). If each agent reasons like this about the other's intentions, they will probable kill themselves in the resulting collisions (mutual defection). Not even a binding agreement will work in chicken, for it violates the very nature of a contest or duel, not to mention that in some cases it may be illegal or immoral to “collude” with the competition to fix the contest. For example, in the case of two companies in a price war it is typically illegal for them to agree to end their game of chicken by fixing prices. And in the case of a personal contest like the beer drinking duel it seems wrong to fool the onlookers with a secrete deal while pretending that a real contest of drinking endurance is going on. In the clash of wills, you will recall, a binding agreement does not work as a rational choice solution because it transforms the original game into another one. Likewise, a binding agreement to mutually cooperate would not mean that the agents achieve a rational choice solution to chicken; instead it would in effect change their game of chicken into a different game.

We should also notice that it does not make much sense to think of chicken as a repeatable game between the same two agents. One time is enough to prove who is superior and who is chicken, and if there is mutual defection (in our example, both hit the cement wall) the agents are typically too damaged to repeat the game. The only case where a second round of chicken seems practically possible is mutual cooperation in the first round. However, in the second round each agent will expect the other to cooperate (chicken-out) based on the first round experience, for each now has gained a reputation of chickening-out, increasing the likelihood that each will stay the course in the second round. You can easily see how this leads to mutual defection and disaster.

If, then, either (R1,C2) or (R2,C1) are equally rational choices and if practical reasoning has here reached a limit about deciding between them, how might a choices happen in chicken? This is where the importance of bluff, deception, and intimidation enter the game as non-rational strategies that can (must?) be used to augment practical reasoning. Each agent in chicken:
(1) secretly must intend to stop and quit the course of action (cooperate) in order to avoid disaster and perhaps destruction, but (2) at the same time must give the appearance to the other agent of being completely and fearlessly willing (indeed, committed) to stay the course so as to force the other agent to back down first. In other words, in chicken you must hide your fears and your very rational prudence, and present an image of strength – an image of daring, imprudence, perhaps

even recklessness, in the face of a potentially damaging challenge. Revealing your intention to give up and to save yourself from harm only serves to reinforce the other agent's feelings of superiority and so will harden the other agent's resolve to outlast you. Bluff and deception are types of information (perhaps misinformation) that will need to be exchanged, so lines of communication must remain open between agents in chicken in order to bring about one of the equilibrium outcomes. That is, the agents must be able to signal each other, or perceive each other's behavior, or somehow have ways to learn about each other's apparent intentions. Let's look at some examples of how an agent, while secretly intending to give up the challenge and accept being the chicken rather than risk destruction, might nevertheless try to communicate the deceptive appearance of a readiness to accept disaster.

Think of the teenage car game of chicken. Suppose one teen arrives at the contest with alcohol and looks drunk or gives the appearance of being high on drugs and acts as if he is ready to die (but was actually sober or not high, just bluffing out of a strong desired to avoid being the chicken). The other teen, seeing this, would be very motivated to stop his car well before the cement wall or cliff, for his rival seems out of his mind and not in a condition to judge danger accurately. Or, suppose one teen yelled over to the other as they were taking off from the starting line that he had disconnected the break system of his car, implying that he would not be the chicken even if it meant his death. Even though this is false, the rival agent could not be sure he wasn't dealing with someone completely irrational and so would be more incline to chicken-out sooner than later.

Take the example above of the two nations in an arms race. How might one nation deceive or bluff or intimidate the other into a decision quit the race, leaving the one nation superior? Suppose one nation spread disinformation that it had plans to produce a new horrifying secret weapon, how would the other nation react? If this attempt at intimidation is believable and if the other nation could not build an equal weapon, don't you think it would be more inclined to give up the arms race and be suckered? Or suppose that one nation announced that it was giving up the arms race (while secretly continuing the military buildup). If it presented this deception well, the other nation might believe it is now superior and slow down its military buildup, allowing the other to pull ahead. How about this ploy: one nation's diplomats tells the other nation's leaders that it would like to discontinue their dangerous arms race, but the decision is really out of its leaders' hands; it has a form of government in which such decisions are completely in the hands of very frightened (and thus irrational) voters who will not vote to stop military buildup in their country no matter what happens until they are assured that you, the other nation, have stopped your buildup first.

Go back and consider the examples of the two college students in a beer-drinking duel, the two watermelon-eating contestants, the two rivals dating the same girl, and the two car companies in a price war. How might an agent in these games of chicken deceive or bluff or intimidate the other agent to give up first? The general strategy, recall, is to appear to the other agent as ready, willing, and able to stay the course to the very end, no matter how damaging that might be; meanwhile, in truth, the bluffing agent does not intend to continue the duel to the point of destruction. Such deception is done as an <u>addition</u> to the methods of practical reasoning, and this addition is needed because practical reasoning discovers two equal equilibrium outcomes and so cannot justify one rational choice solution. Yet each agent desires to achieve just one equilibrium outcome, namely, the one in which he achieves superiority and makes the other agent the chicken; that's the goal.

Whatever you might recommend in these examples of chicken as a bluff, deception, or intimidation, it must be done <u>convincingly</u>, for if the other agent senses any hint of weakness or catches on that the deception is nothing but deception, a trick and not a serious intention to stay the course no matter what, it will only serve to increase the other agent's strength and severely weaken the bluffing agent's position. The problem, of course, is that both agents in chicken, under our general assumption that they are equally motivated to achieve the goal, are trying to deceive, bluff, and intimidate the other into giving up first. (R1,C2) and (R2,C1) are equally rational choices; what will tip the balance giving the goal to one and the sucker's payoff to the other is the skill with which each side practices this particular form of deception. If they are equally skilled and equally willing to stay the course (which is unlikely), the mutual defection payoff will result. If they are equally willing to chicken-out (more likely, given they are equally rational and that this is common knowledge), they tie with the mutual cooperation payoff and an equal loss of the goal.

One of the most convincing ways for an agent to fake a serious determination to stay the course is to get the other agent to believe that the decision to cooperate (to chicken-out) is no longer within the agent's power to make and so the agent is irrevocably – irreversibly – committed to staying the course to its very end. Recall that once an agent in chicken believes this about the other agent, the rational choice for that agent is to cooperate and be the chicken, the sooner the better the more danger there is to the challenge. Two ways that achieve the required level of convincing are (1) to appear irrational or (2) to appear in a decisionally weakened condition. So, for example, one teen shows up pretending to be drunk or high on drugs and so pretends to be out of his mind and crazy. To illustrate appearing to be in a decisionally weakened condition, we imagined one teen announced to the other that he has disabled the breaks and so couldn't stop even if he wanted to, and we pictured the leaders of a nation making the claim that their hands are tied (for the decision rests with frightened voters over whom they have no control). As you can imagine, it is very intimidating for a rational agent to believe that he is in a game of chicken with another agent who might be irrational or whose back is against the wall and thus lacks the decisional ability to cooperate no matter how much skillful deception and bluff is used on him. Look at the matrix that represents this situation. Suppose for the moment that Row successfully deceived Col into believing that Row was either so irrational or had his decision-making ability so out of his hands that he was no longer able to choose to cooperate. Col's decision problem would then be:

		Col	
		C1: cooperate (stop first)	C2: defect (stop last)
Row:	R2: defect (stop last)	4, 2 (10, -9)	1, 1 (-10, -10)

Col would have to choose to be the chicken in such a situation, for the only alternative would appear to be mutual ruin. The result, then, of Row's skill at deception and intimidation is the (R2,C1) Nash equilibrium outcome giving Row the goal.

If both agents are equally rational, and equally skilled at the art of bluffing, the expected result is a tie: mutual cooperation. Neither is the chicken, and neither is superior. Each has successfully deceived the other into choosing the cooperative option. For very dangerous games of chicken like that of two nations in an arms race, or two college students in a beer-drinking contest, or two auto companies in a price war – the aspect of danger perhaps best illustrated in the teenage car-duel version were lives are at stake – it seems best for all stakeholders to hope that each side is equally rational and in addition equally skilled in the art of deception. If they are equally poor in the art of deception, each agent seeing through the other's attempt to deceive, mutual defection (destruction) is the likely outcome.

11.5 Summary: equilibrium selection problem as a failure of practical reasoning

From the analysis and evaluation of the above two potentially cooperative games, the clash of wills and chicken, it seems we must accept the failure of practical reasoning to justify a single pair of options, one for Row and the other for Col, as the rational choice. The same methods and principles of practical reasoning that proved so powerful in arriving at rational choice solutions in competitive decision problems seem inadequate when it comes to some non-zero sum games that seem especially insightful for understanding a wide range of human cooperative decision making. Let's take stock by briefly comparing the three games covered in this chapter before considering two more troubling cooperative games in the next chapter.

Harmony, you will recall, was our standard. In this game, agents cooperate as the rational choice justified by methods of practical reasoning without any non-rational "outside" assistance in the form of affection or good will toward one another, or a third-party enforcer such as government, or religious duty or moral obligation. This is not what happens, however, in the other two potentially cooperative games. The strong clash of wills and chicken are similar in that both suffer from "too much." They each have one-too-many rational choice solutions; that is, they each have more than one equally rational Nash equilibrium points, and there seems to be no practical reasoning way for agents to coordinate decisions one way or the other. As far as practical reasoning goes, there is no single rational choice solution. Instead, unsatisfying (at least to those who believe in the central role of human reasoning in making good decisions) recourses to non-rational "outside" decisional help must take place for one or both agents to maximally achieve the goal (e.g., flipping a coin or skillful use of deception). In these two important patterns of human interaction, then, practical reasoning leaves agents with an **equilibrium selection problem**, threatening the prospects for cooperation with a stalemate or deadlock.

This failure of practical reasoning links to a larger issue. Recall the brief description in Chapter 8.4 of the rationalist/behaviorist debate concerning maximizing verses satisficing. The broad issue was this: how should the principles and standards of rational choice theory be understood? Do they have the status of rational norms that we should try to conform to in our decision making (as the rationalists argue), or are they open to empirical testing and consequently open to being adjusted to conform to the reality of how real agents make decisions (as the behaviorists argue)? For the rationalist, rational choice theory offers standards and methods discovered and justified on rational grounds independent of how people actually make decisions (in the same way that, for example, logic discovers and justifies standards of valid inferences independent of how people actually make inferences). These standards and methods are the norms by which actual choices can rightfully be judged good (rational) or bad (irrational) decision making. If it turns out that most people most of the time actually make irrational choices, the problem does not lie with the theory, the problem lies with the way people go about making their real decisions. That is, the reasoning leading to their choices is flawed, and so *people* should change their reasoning to better conform to the methods and norms of rational choice.

The behaviorist position, on the other hand, is that the theory of rational choice, if it is to apply to real human decision making, must be based in reality, not in unrealistic ideals. The principles of practical rationality have (or should have) empirical status, which means that they must be verified or falsified on the basis of observing how people factually make decisions. If we find that enough people don't keep to these principles and methods in how they go about making decisions, for example if we find that people are in fact satisficers rather than maximizers, it is wrong to judge them irrational; rather, this shows that the theory in question is false and unrealistic, and *it* must be changed to better fit the facts.

This controversy is not limited to the issue of maximizing verses satisficing as methods of practical reasoning. It gains deeper and wider importance for understanding rational choice in light of the failure of practical reasoning (as guided by rational choice principles) in the cases of the clash of wills and chicken. This failure appears to weaken the rationalist position and strengthen the case for the behaviorist that rational choice theory is seriously limited and "out of touch" with reality in presenting norms of how decisions ought to be made.

EXERCISE:

1) Identify the potentially cooperative games (harmony, strong clash of wills, weak clash of wills, or chicken) represented by the following matrices. (See if you can recognize the game from the matrix by trying to put yourself in each of the agent's shoes, rather than mechanically matching up the matrix with the forms in the chapter.) Which are symmetrical and which asymmetrical, and if asymmetrical which are Row-biased and which Col-biased?

a) example: Col

Row	C	D
C	4, 4	3, 8
D	8, 3	-5, -5

Chicken, symmetrical

b) Col

Row	C	D
C	-2, -3	-6, -5
D	-3, -7	-8, -9

c) Col

Row	C	D
C	.5, 1	7, 10
D	10, 7	1, 4

d) Col

Row	C	D
C	-10, -9	5, 15
D	20, 5	-1, -3

e) Col

Row	C	D
C	0, 0	-2, -6
D	-6, -2	-3, -6

f) Col

Row	C	D
C	7, 8	0, 10
D	9, 1	-2, -1

g) Col

Row	C	D
C	-5, -4	-6, -1
D	-1, -6	-7, -8

h) Col

Row	C	D
C	-3, -6	2, 8
D	11, 2	-1, 0

i) Col

Row	C	D
C	5, 10	-10, 1
D	1, -10	-15, -15

2) Here are four scenarios. (1) Frame each as one of the potentially cooperative game covered in this chapter, (2) identify the game by name, and (3) say why the appropriate principles of practical reasoning justify or fail to justify a rational choice solution.

a) Sally and Joe are discussing their wedding plans. His family is insisting on a traditional wedding: religious, a huge family gathering, and a grand reception – the works! Joe would like to please his family, and in addition has his eye on a long honeymoon at an up-scale resort hotel. Sally is very uncomfortable with all this; she wants a small civil wedding with minimal costs, and a honeymoon that involves hiking and camping in a wilderness area. Her idea is to save the money that a big wedding and an expensive honeymoon would cost and, with their savings, put it toward a down payment on a house. Neither Sally nor Joe is softening; in fact, because each feels so secure and trusting in their relationship, they are each hardening their position about the kind of wedding that they will have. A "happy medium" is out of the question; neither can warm up to a wedding that would both disappoint Joe's family expectations and yet still cost a good chunk of their savings and wedding gifts.

b) Abby and Beth have formed a last minute study group for final exams in a notoriously difficult course in which neither is doing very well. Each has promised to prepare a different hard unit of material well and teach it to the other. Neither has time to study both units well enough to pass. This will take time away from their active social lives, but there is a good chance that knowing both units will allow each to pass the course with a C grade, whereas neither can achieve more than a minimally passing D grade knowing just one unit, and will earn a failing grade if neither unit is prepared. The problem is that Abby and Beth can go to a pre-finals party that everyone knows is one of the best of the school year. Each feels very tempted by the party; but they can't both party and prepare the unit. Even though Abby and Beth are each unsure what the other will decide to do (prepare the unit or go to the party), each values passing the course with the best possible grade and keeping her word more than enjoying the popular pre-finals party.

c) The international scene is heating up. A small country (R), long suspected of harboring terrorists, is now suspected of developing nuclear weapons, in violation of a United Nations agreement its previous leaders signed onto. A neighboring country (C), likewise a signatory to the UN non-nuclear weapons agreement, has a troubled history with R. C will not allow R to be the only country in the region with nuclear weapons, and has issued a challenge: R must give up its nuclear weapons development or C will launch a military strike against R's manufacturing facilities, and in addition will begin its own nuclear weapons development. R has responded that because of C's hostility and threats, it will continue and even speed up developing nuclear weapons as the best protection against C's possible aggression. Each side would like to be the only country in the region with "defensive" nuclear weapons. Each side believes that if both have nuclear weapons it will likely result in a war of total destruction. But the current status of neither having such weapons is better than being without them while the "other side" develops them.

d) Maron and Raina are twins in high school who form the more important half of a locally popular rock band "Total Deconstruction": Maron is the band's lead guitar, Raina plays bass guitar. The twins are responsible for creating and arranging new material, and scheduling practice sessions. The band has recently landed an important gig, and the band has a good chance of getting air-time at a nearby University radio station if the gig is a success. They both want success very much, but Maron believes that the band will achieve most success in their gig by sticking with material and arrangements that the band knows well and has performed hundreds of times. Raina, however, has some rough ideas for new material, requiring a lot of work and practice to pull it off; she believes that the band sounds "stale" doing familiar material and will be a big success with the excitement and challenge of new material and arrangements. Neither twin is ready to give in to the other; much ride on the success of this gig, it's the band's big chance! They can't do both the old and the new material; that would be a terrible performance, equal to doing neither and calling the gig off. In the heat of their argument, both twins do agree that there is more risk of failure in Raina's idea for the gig than there is in Maron's.

The Sources and Suggested Readings section found in Chapter 12 covers both Chapters 11 and 12.