

Notes on David Gauthier, *Morals by Agreement*.

Dick Arneson For Philosophy 160

The notes summarize Gauthier's ideas, including some pieces of his account not discussed in our assigned Gauthier readings. Critical comments and questions are enclosed in square brackets, like this--- [Gauthier's ideas are bogus.]

Let's say that egoism is the view that on every occasion of choice, one ought always to do whatever act would maximize the satisfaction of one's own advantage. (Egoism does not necessarily recommend that one should always act selfishly and pay no heed to the interests of others. Sometimes helping others encourages them to help you in return and thus works to maximize your advantage in the long run.)

David Gauthier's view is close to egoism. He holds that the sole rational goal for each person is maximizing the satisfaction of her own interests. Here one's interests are fixed by one's basic (noninstrumental) preferences or desires. If you happen to desire the good of others in some ways or to some degree, then your interests include achieving the good of others in some ways or to some degree. But the mere fact that another person would benefit from getting some good does not per se give you any reason to bring about the person's getting that good.

Gauthier thinks that practical reason so understood justifies moral constraints. He says, "We shall defend the traditional conception of morality as a rational constraint on the pursuit of individual interest" (p. 2). Example: If morality as traditionally conceived includes the rule that one should not steal other people's property, this rule is a constraint on the pursuit of my own interests. Morality tells me not to steal other people's property even when doing so would be the best way available to advance my interests. How can this be? If what is rational is maximizing the satisfaction of one's interest and morality constrains the pursuit of one's own interest, how can acceptance of moral constraint be rational? Gauthier asserts, "rational constraints on the pursuit of interest have themselves a foundation in the interest they constrain. Duty overrides advantage, but the acceptance of duty is truly advantageous" (p. 2).

Here's a story that illustrates what Gauthier has in mind. Suppose that if you are a nice, cooperative person, one who keeps her agreements, you will be recognized as such by other persons and admitted by them into mutually profitable cooperative arrangements. If you are not a nice, cooperative person, but instead prone not to keep your agreements, you will be recognized as such by other persons and shunned by them. They will avoid interaction with you. If the social world you face has these features, then it is easy to see that being disposed to keep one's agreements will sometimes lead you to do disadvantageous acts—acts that are really disadvantageous, disadvantageous to you even in the long run, not merely seemingly disadvantageous or disadvantageous-in-the-short-run-but-advantageous-in-the-long-run. However, being disposed this way will still lead you to be better off if the disadvantages of keeping your word are outweighed by the advantages of being recognized as a trustworthy person and being admitted into mutually profitable associations. Having the character trait of being disposed to keep one's promises can be advantageous even if the act of keeping one's promises is sometimes truly disadvantageous. Example: suppose Tom and Sally are farmers growing crops on adjacent fields. Tom's crop ripens first, Sally's a few weeks later. Both will be better off if they Sally helps Tom harvest his crop and then Tom helps Sally harvest her crop. So a mutually advantageous deal is possible. But this is threatened, if Tom acts according to egoism on each occasion of choice. Having already been helped by Sally, he has nothing to gain from helping her later. So he won't. Foreseeing this, Sally will not help Tom first. If Tom had disposed himself to be what Gauthier calls a constrained maximizer, specifically one who pursues his own interests except that he keeps his promises, then if Sally recognizes this is so, she will be willing to make a deal with Tom and be willing to help him first, foreseeing he will do his part of the bargain later, when her crop ripens. Note: In this situation, Tom's helping Sally to harvest her crop is really and not merely seemingly disadvantageous for him.

The story just told about keeping promises and agreements can also be told about other moral constraints such as truth-telling or refraining from stealing other people's property.

Of course, whether disposing oneself to accept moral constraints is likely to be advantageous for oneself depends on the specific constraints being considered and the social environment in which one is going to live. The moral trait of being charitable to those in desperate need might not be advantageous in the way being considered, if being recognized as a generous and charitable person would not help one get admitted to valuable cooperative schemes. Also, suppose one lives in a social environment in which everyone else or almost everybody else is strongly disposed to lie and break their promises whenever doing so looks to be advantageous for the promise-breaker. Disposing oneself to be honest and a faithful promise-keeper might be disadvantageous, not advantageous, in such an environment. For more on these complications, see Gauthier, chapter 6.

[Question: Gauthier seems to assume that the constraints it is rational to accept as advantageous to oneself will be nice, impartial constraints that are recognizable as ordinary moral requirements (keep your promises, tell the truth, respect private property, return good for good and evil for evil, and so on). Is this so? Maybe in some social environments, it will be advantageous to form the disposition to exploit and cheat the weak and vulnerable, play fair with those who are about as powerful as you are, and kowtow and defer to the powerful and even accept considerable abuse from them. More on this when we turn eventually to discussion of chapter 6. Another possibility: maybe in some social environments it would be advantageous to dispose oneself to be a bullying person who makes threats and fulfills the threats he makes even when doing so is disadvantageous to himself. If I am known to be one who will kill when I say "your money or your life" even when killing is risky and costly to me, that threat will be effective, and people will have strong reason to comply with my threats. In this way I am a type of what Gauthier calls a constrained rather than an unconstrained maximizer, but the constraints I accept are nasty not nice. Another possibility: instead of accepting the principle "don't steal" one might find it more advantageous to accept a qualified principle such as "don't steal other people's property unless you are faced with a windfall profit opportunity—a situation in which you can gain a huge benefit by theft and the chances of getting caught are slim or zero. Again, the constraints it would be advantageous to accept in many likely environments seem likely to turn out to be nasty not nice.]

[Thinking about these possibilities, we may be led to question the starting premise from which they are derived—that rationality in choice of action involves maximization of one's own preferences. Why accept this?]

In chapter one, Gauthier introduces several ideas that shape his account of morals by agreement. The phrase "morals by agreement" conveys the idea that moral constraints that are rational to accept are those that would hypothetically be agreed to by people each of whom is seeking to maximize the satisfaction of his own interests, when faced with reasonably favorable environments and when each knows her own circumstances. The key ideas he mentions are: the idea of a "morally free zone" in which constraints of morality have no place, the idea of rational bargaining when there are gains to be had from cooperative interaction, and the idea of constrained maximization. I'll run through each of these briefly in turn. Then I'll summarize Gauthier's views about reason and value (chapter 2).

1. The market as a morally free zone. It will help to introduce some Gauthier terminology. He distinguishes parametric and strategic choice. You choose parametrically when all of the circumstances that bear on your choice are fixed (maybe probabilistically fixed) and nothing is unsettled except your choice and what follows causally from your choice. You choose strategically when you are interacting with another person or persons. In interaction, what other persons are choosing has a bearing on what it is rational for you to choose, and what you are choosing has a bearing on what it is rational for the others to choose. As Gauthier sees it, moral constraints have a role to play in contexts of strategic choice, when there are mutual gains to be had from cooperation. When you choose parametrically, if it is known with certainty what

outcome will ensue, given any choice of action you might take, rationality dictates picking the action that yields the outcome that is most advantageous to you. If you don't know for certain what will happen following your choice, but can associate, with every action you might now choose, each of the possible outcomes that might ensue given that choice, the value to you of each outcome if it occurs, and the probability the outcome will occur given your choice, then rationality dictates that you ought to choose the action with the highest expected utility (see Gauthier's account of expected utility maximization in pp. 42-44).

When people are acting in a perfectly competitive market, there are many buyers and many sellers of each good being offered for sale. (There are further conditions required for perfect competition.) Prices are fixed by the ensemble of supply and demand conditions, and no individual can affect the prices of goods. Each then chooses to maximize her utility by producing goods or purchasing goods or both. Choice here is parametric. Hence, according to Gauthier, there is no room for moral constraints. Each acts as an unconstrained maximizer in this setting. Under standard conditions, buying and selling on a perfectly competitive market leads to an outcome that is Pareto-optimal—no one can be made better off unless someone else is made worse off. So moral constraints imposed on the perfect competition setting will not make some better off except by making another worse off. Again, according to Gauthier, there is no sensible role for moral constraints to here.

[Contrary to what Gauthier says, a perfectly competitive market seems to embody moral constraint. Rather than pay you a dollar for the onions you are selling, I might instead bash you and take your onions for free. The market would then no longer be perfectly competitive, but nothing in the idea of perfect competition rules out the possibility that someone could improve her situation by theft, robbery, extortion, assault, or some other predatory activity. Gauthier obliquely addresses this point in another chapter. He affirms that rational individuals prior to market activity will refrain from predatory activity and indeed from any activity that imposes costs on others (renders another person worse off than if you weren't present at all), at least if they anticipate the formation of market activity. Why so? Gauthier thinks rational individuals would exclude people who impose the costs of their behavior onto others from bringing these gains with them as their endowment at the start of market activity. Instead rational individuals will exclude predators and polluters from participation in the market. So to gain from such participation, rational individuals have incentive to refrain from predation and pollution activities prior to market activity (and one supposes, in the course of market activity as well). The problem with this gambit is that I do not see why rational self-interested individuals would necessarily and always assume the costs of excluding predators and polluters from market exchange. These people might bring large endowments to the market, and there might be large gains from trading with them. Why forego those gains?]

[Notice that from moral standpoints other than Gauthier's, moral norms might seem appropriate regulators of people's behavior when they are in a perfectly competitive setting. Example: there is a highly competitive market involving purchase of housecleaning services in San Diego—many buyers, many sellers, a competitive market price for these services. But why does this set of facts insulate me from moral criticism of the payment I make to Carmen, who cleans my house? The competitive market price does not coerce me. Suppose the market price is ten dollars per hour. She might be willing to do the work for anything over five dollars per hour and I would be willing to pay as much as \$75 per hour and still benefit. So nothing blocks me from paying her say \$60 per hour and maybe I ought to do that. If I do, the market is no longer perfectly competitive, but so what?]

[Even if the criticisms of Gauthier that I am gesturing at here are sound, they would not impugn Gauthier's main claim: that in some social environments people disposed to abide by certain moral constraints (the people he calls constrained maximizers) will do better for themselves than those who on each occasion of choice act rationally to advance their interests (the people he calls straightforward maximizers).]

2. Rational bargaining. Suppose that Sarah and Samantha can form a partnership and do better working together than either could do on her own. To form the partnership, they must agree on how to split the gains from their cooperative activity. If they cannot agree and do not form the partnership, they lose the potential gains. Gauthier sees a general issue here. When people can interact in ways that produce gain, rational agreement requires rational resolution of the bargaining problem of how to divide the gains from cooperation. (Even if people are engaged in hostile interaction, cooperative gains may still be possible. Two countries may be at war, and both sides may have higher expected utility if they can agree to forego the use of biological warfare agents. Here again there is a bargaining problem.)

You might think there is no general rational principle that stipulates what should occur when people bargain in these settings. The results of bargaining will differ from case to case depending on particular features of the bargaining parties, and there is no uniquely rational answer to the question, should I hold out for a greater share of the cooperative gains even though this hold-out increases the risk that no agreement will be reached. Gauthier thinks there is a unique general of the bargaining problem, This is the principle of minimax relative concession.

Gauthier thinks that the idea of interpersonal comparisons of utility or well-being lack content. We can't compare how much utility Andre gets from an apple and how much utility Alessandra would get from eating the apple. He does think that intrapersonal cardinal comparisons of utility or well-being are possible. This just means that Ted can determine whether he would now like playing video games more than playing golf, and how much more he would prefer playing the one game than the other.

Now go back to Sarah and Samantha. Each compares how much utility she would gain if she collected all of the gains from cooperation for herself to the utility she gets in the absence of cooperation. What she gets from bargaining minus the maximum she claims (giving all gains to herself) is her concession. Her relative concession is the ratio of the absolute magnitude of her concession to the absolute magnitude of complete concession (in complete concession, she gets the utility she would get from no agreement and no cooperation at all). Gauthier says the unique solution to any bargaining problem is that the bargainers should agree on a deal so that the relative concession of the person whose relative concession is greatest is as small as possible. This is the principle of minimax relative concession (make the largest concession as small as possible). This idea is stated above for two-person bargaining but can be generalized to negotiations involving many bargainers.

Gauthier thinks that rational persons will hold out for minimax relative concession (when they can get it) and will not insist on more. They will not cooperate with others except on terms that are close to minimax relative concession. (Gauthier backs away from this stringent rule, saying if you can gain from cooperation but cannot get minimax relative concession, you might have to accept being exploited if that is the best you can do but you will not be a willing cooperator. When you have the ability to hold out for minimax relative concession, you will do so.

[Comment: I think that self-interested bargainers will hold out for what they can get. If they can get more than Gauthier's favored principle allows them, they will press for more. If they must choose between getting less than what Gauthier's principle would allow them and losing out, they will accept less. Gauthier's minimax relative concession principle is in effect his ideal of fair bargaining, but I do not see why Gauthier persons—persons who are striving to maximize the satisfaction of their preferences—will care about fairness except insofar as a fairness norm happens to work as a tool or weapon to help them get more for themselves.]

3. Constrained maximization. Suppose Sarah and Samantha make a deal according to Gauthier's favored resolution of the bargaining problem or some other solution (or maybe they just haggle and accept some result, not on any principles basis). Next question: Will the two carry through on the bargain they have made? Bargains like this will be just words unless each

expects the other to fulfill her part of the bargain. To simplify matters, suppose it is clear that Sarah could cheat Samantha and vice-versa. They are setting up a business; Sarah manages sales and can take money from the cash box. Samantha deals with suppliers and the assembly line and can channel material away from the firm and into her own hands and sell firm assets for her own profit. So the partnership is mutually profitable but only if each is trustworthy.

In this setting, a straightforward maximizer, who acts to further her own interest on each occasion of choice, cannot get the gains of partnership. The deal will fall through. One will not trust a straightforward maximizer (SM). Note also that an unconditional cooperator, who will behave honestly and do the “right” thing whatever her partner is doing, will predictably get exploited by the other. Gauthier thinks that the rational individual will dispose herself to be a constrained maximizer—she cooperates with those who cooperate with her. (We’ll say more about constrained maximizers (CMs) when we look at Gauthier, chapter 6, next class.

The CM steers between two extremes. She does not exploit others who are cooperating with her in ways that give her a payoff close to minimax relative concession. In the Sarah and Samantha case, she will not steal money from the cash box even if she knows she could get away with this theft from her partner. On the other side, the CM does not cooperate unconditionally. If she knows or suspects her partner is a wolf (a SM), disposed to cheat her, she does not cooperate with such a person. Facing SMs, the constrained maximizer behaves as another SM.

Gauthier uses single-play Prisoner’s Dilemma to illustrate the key attributes of a CM. She cooperates in single-play PD if she has good reason to believe the person she is interacting with is cooperating as well, and otherwise does not cooperate. (This idea gets explained in the notes for next Monday’s class.) Notice that how the CM behaves depends on her recognition of the character of those she interacts with—are they predatory or cooperative? So rationality requires that she have or develop such recognitional ability. Also, it is not unconditionally rational to dispose oneself to be a constrained maximizer, a self-interested cooperator. The rationality of disposing oneself to be CM (forming one’s character so one makes CM choices) depends on the proportion of cooperative and noncooperative people she expects to meet and interact with in the social environment she will inhabit over the course of her life.

Putting the pieces together, Gauthier holds that his account shows how according to a rational morality, “Duty overrides advantage, but the acceptance of duty is truly advantageous.”

Chapter two—Choice: Reason and Value. An interesting feature of Gauthier’s account is his unrelenting subjectivism about value. The aim of a rational agent according to Gauthier is the maximal satisfaction of her preferences. This phrase needs some elaboration and interpretation.

Preferences or desires are attitudes one has. They explain choice. But not just any choice on the basis of one’s preferences counts as rational. There are conditions one’s preferences must satisfy, if they are to provide a rational guiding basis for choice.

1. One’s preferences must be coherent. If you prefer A to B, and B to C, you must also prefer A to C. If not, your preferences are intransitive, and fail to provide a coherent basis for choice. (Gauthier introduces further formal consistency requirements on preference, which we here leave aside.)
2. Your preferences must be expressed consistently in verbal and choice behavior. This condition is satisfied when you both sincerely say you prefer rock music to opera and when given a choice, you pick rock music over opera (other things being equal). To the extent that your sincere verbal expressions of attitude and your choice behavior do not cohere, your preferences are a defective basis for choice.

3. Your preferences must be considered and based on experience. Your preference fails to be considered if you form it without reflection or deliberation. Preferences formed by reflection and deliberation are considered preferences. (An alternative idea would be that your preferences are ill-considered just in case they would be altered if you were to engage in reflection and deliberation concerning them. I have not reflected carefully about my preference for gooey desserts, and would not know how to begin, but this preference might be such that if I were to deliberate about it, it would not vanish or alter.) An inexperienced preference is one not based on relevant experience of the thing one wants. An experienced preference for wheat beer is one formed by sampling various types of beer.

**

Choices that satisfy these three conditions are an adequate basis for choice. The rational object of choice is the maximal satisfaction of your preferences, to the degree they satisfy these three conditions.

**

---What should we say about choices that go awry due to false factual beliefs held by the choosing person? Gauthier gives the example of Queen Gertrude, who drank from the poisoned goblet, thinking that the goblet contained wine not poison.

The problem here does not impugn Gertrude's preferences, according to Gauthier. From what we are told, her preferences might be fine, not defective at all. Given her false factual belief on a matter that was material for choice, her choices do not adequately reflect her preferences. She does not want to die from poison, she wants to drink a glass of wine that is to her taste and gives her a pleasant buzz.

---What should we say about a person whose coherent, considered, based-on-experience preferences are unstable over time, alter with passage of time? Example: Carlos prefers baseball to soccer right now. A year from now, he will come to prefer soccer over baseball. What preferences here should be the basis for Carlos's choice? He might or might not know or suspect that his current preferences will alter over time in a certain direction; just suppose he does know.

Gauthier identifies rationality with satisfying the consistent, considered, experience-based preferences one has NOW. Preferences you used to have and have no longer are not relevant to choice now. Nor are preferences you lack now but will have in future (unless you have a preference now that your future preferences, all of them or specific ones, be satisfied).

Since Carlos now prefers playing baseball to playing soccer, it makes sense for him to bring it about that he will have a baseball, not a soccer ball, in his possession a year from now (when, as he knows now, he will prefer soccer over baseball), if that will induce him then to play baseball not soccer. Carlos acts now to bring about satisfaction later of his present desire to play baseball later. The fact that a future stage of himself will have a different preference in this matter is according to Gauthier a good basis for his choice later, when his preference changes, but not now.)

Gauthier is opposed to impartiality across persons. What is rational is for me to strive to satisfy my preferences and you to strive to satisfy yours, not for either of us to strive to satisfy people's preferences impartially, no matter whose preferences they are. In the same spirit, Gauthier rejects the ordinary conception of prudence, according to which one ought to treat impartially all the times of one's life, acting to maximize the satisfaction of all preferences one has throughout one's life. The prudent agent accept the slight annoyance of tooth-brushing now to avoid the greater pain of tooth decay and infection later in life. The Gauthier-rational agent, if she cares about avoiding annoyance now but does not care what happens to future stages of herself, now prefers to avoid the annoyance of tooth-brushing at a cost of later tooth decay, and acts on that basis. The person later will maybe have opposed preferences, and regret the failure earlier to have brushed one's teeth, but that is according to Gauthier a problem of life not of rational choice theory.