Cambridge Books Online

http://ebooks.cambridge.org/
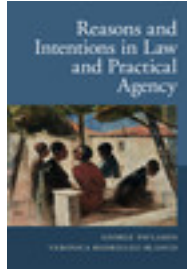

Reasons and Intentions in Law and Practical Agency

Edited by George Pavlakos, Veronica Rodriguez-Blanco

Book DOI: http://dx.doi.org/10.1017/CBO9781107707573

Online ISBN: 9781107707573

Hardback ISBN: 9781107070721

Paperback ISBN: 9781107642928

Chapter

2 - Acting and satisficing pp. 31-51

Chapter DOI: http://dx.doi.org/10.1017/CBO9781107707573.004

Cambridge University Press

2

Acting and satisficing

SERGIO TENENBAUM

I am searching for new shoes. I go to the first store, look at the selection and prices, move to the second store, find a pair a bit cheaper and nicer than my favorite in the first store, and buy it. My friend comes to me and says: 'This might come as a surprise to you, but there are many more shoe stores in Toronto; in fact, dozens of them. The chances that you'd find an even nicer pair for an even cheaper price in one of these stores are very high. Fortunately, you still have time to return these shoes and continue searching; I expect you will do much better.' Sensible words indeed. But I reply: "Thanks for your insightful remarks, but these shoes are *good enough*. I shall keep them." My response seems perfectly rational, and such examples seem to speak for the correctness of satisficing as a standard of rational choice. According to a general satisficing theory of rationality, an agent is rational if (and only if) her choices are good enough, and thus an agent's choice can be rational even if the agent recognizes that there is another choice that is (or is expected to be) better.[1] My expectation of finding better shoes were I to continue my shopping adventures does not show that my decision to buy the second-store pair was unreasonable; it is perfectly rational for me to turn a deaf ear to my friend's advice.

Satisficing is thus supposed to be an intuitive alternative to maximizing theories of rationality which would require me never to choose an option if there is any other option I can choose that has a higher expected value or utility.[2] Almost no one doubts that, given normal background

[1] In some ways of understanding this definition, this would be a rather strong form of satisficing. I will refine the view as I continue.

[2] It does not matter to my purposes whether "expected value" in our understanding of maximization is a function of the agent's preference or a function of belief or some belief-like attitude towards objective value. I will alternate between these understandings in the chapter.

assumptions, it is perfectly rational for me to keep my new shoes. Yet, it is dubious whether such examples really pose a challenge to a maximizing conception of rationality. Maximizers will easily come up with ways of representing the intuitively permissible course of action as a maximizing action. In our case, maximizers could say that the reasonableness of stopping our search short of the expected best shoes is due to the overall expected gain in utility rather than being due to some supposedly rational attitude of frugality towards value as such. After all, if I were to keep shopping for shoes I would need to forego meeting my friend for coffee, have unpleasant interactions with other sales clerks, feel the agony of not knowing what my full outfit will look like for a few more hours, etc.

However, advocates of satisficing do not merely rely on examples. They often try to provide some kind of rationale that together with such examples would show a satisficing theory of rationality to be a well-motivated theory. Defenders of satisficing mainly rely on the limits of human cognitive faculties, special features of human agency, or certain types of values in generating such rationales.[3] But these theoretical defenses of satisficing face challenges of their own. In this chapter, I present in broad outlines the problems that satisficers typically face in defending their views. Then I present a very different kind of rationale for satisficing – one that does not depend on any of these limits, or peculiar features, of human agency. Rather, the defense of satisficing I propose depends only on very general features of practical rationality. For this very reason this view is not prone to any of the typical difficulties encountered in various attempts to make sense of satisficing. In particular, I argue that once we move our attention to rationality in *action* proper (rather than desires or preferences), especially long-term actions with indeterminate ends, we see that satisficing has a very important role to play in our understanding of rational agency.

A dilemma for satisficing

Most plausible versions of satisficing rules *permit* rather than obligate; no one claims that you ought to forego a (known) maximizing option

---

[3] Many proponents of satisficing argue in favor of ethical satisficing as a form of conse-quentialism that can survive the "demandingness objection." I am not concerned with such views. I am only concerned with satisficing as a general decision rule. For a survey and criticism of various possible versions of ethical satisficing, see B. Bradley, "Against Satisficing Consequentialism," *Utilitas* 18 (2) (2006), 97–108.

because a suboptimal option clears a certain acceptable threshold. Advocates of satisficing generally claim only that it is not irrational to do so.[4] So here are two possible satisficing rules:

> **General Satisficing Rule** (GSR): For any choice set C, there is an option O^t such that for every option O at least as good as O^t, if O is in C you may choose O (even if there is another option O^+ that is better than O in C) simply because O is at least as good as O^t.[5]

> **Restricted Satisficing Rule** (RSR): In some choice situations (you may) choose option O, even if there is an option O^+ available such that O^+ is better than O simply because in such a context one is allowed to choose any option that is at least as good as O^t and O is such an option.

GSR and RSR do not cover all the possible versions of satisficing; for instance, they eschew comparative versions in favor of absolute ones.[6] However, sticking to these principles allows us to avoid various complications that are ultimately irrelevant to the argument below. It is worth focusing on the explanatory clause to understand the difference between the two rules. GSR claims that rationality allows for the existence of basic "value thresholds"; that is, for the rational pursuit of certain actions in face of better options simply because these actions are "good enough." RSR, on the other hand, is not committed to the view that actions can be

---

[4] Not all versions of satisficing formulate the rule as a permission. P. Pettit, "Satisficing Consequentialism," *Proceedings of the Aristotelian Society*, supplementary vol. 58 (1984), 139–76, for instance, proposes a rule that tells us to choose, roughly, the first choice above a certain threshold. However, in proposing a dilemma for satisficing, it is worth noting that it arises even for weaker views. It will also be clear in what follows that in the positive proposal I defend, the satisficing rule is indeed the only possible rational rule to follow, not because you are *required* not to maximize in these contexts, but because it is not *possible* to maximize in these contexts.

[5] Neither O nor O^t needs to be in the set, so it is compatible with GSR that there are certain choices that you need to maximize, since you cannot reach the "good enough" threshold. A more precise definition would specify that O^t cannot be defined in terms of the maximizing action in each choice set or the maximal point in the agent's preference ordering, so that the maximizing rule does not turn out to be an instance of satisficing rules. An obvious way to do this would be to stipulate that O^t is never the maximal point in the agent's preference ordering and that replacing an option from C with another of higher utility does not change the value of O^t. Alternatively one could change the order of the quantifiers and stipulate that there is no unique value for O^t (other than the maximal point) across all choice sets. But it is easier to work with the simpler formula, so I will simply assume that one of these (or similar) stipulations hold.

[6] See T. Hurka, "Two Kinds of Satisficing," *Philosophical Studies* 59 (1) (1990), 107–11, for the distinction. Nothing significant in our argument would be affected by looking into the comparative version.

made rational simply by being good enough; rather, RSR claims that particular features of a choice situation justify a permission to choose an option that is suboptimal. It is easier to see the difference between the two principles if we think of items in a choice set ranked in terms of a certain value. So, let us say C's members are all aesthetic experiences (let us say that this is a choice among contemplating various paintings). Now suppose that GSR determines for this case that any aesthetic experience that is as good as contemplating painting P is "good enough." Then, according to GSR, whenever I choose to contemplate a painting that is at least as good as P I am rational, even if I could have as easily chosen to contemplate a better painting instead.[7]

RSR, on the other hand, does not imply anything nearly as general. It will, perhaps, allow me to choose to contemplate a painting that is "good enough," if it would be difficult to deliberate about the various options on this occasion, or if I am under time pressure and I tend to be a terrible reasoner in pressure situations, etc., but it would not give me a blanket permission to choose a less valuable option over a certain threshold.

With this distinction in mind, we can put forward a basic dilemma for arguments for satisficing. GSR seems extremely implausible. It is hard to get behind the idea, for instance, that, in the absence of any competing considerations, I could choose the (equally priced, equally fattening, etc.) WORSE QUALITY ice-cream over the PREMIUM QUALITY one simply because WORSE QUALITY is good enough. On the other hand, if we do specify certain features of the context, satisficing becomes more plausible. If, as I am about to leave the store after having bought the WORSE QUALITY ice-cream, I see that a new shipment of PREMIUM QUALITY has arrived, it seems rationally permissible that I forego the opportunity to take advantage of their "Easy Refund, No Questions Asked" policy.

Similarly, to use one of the classic examples in support of satisficing, if I am trying to sell my house, it seems permissible that I settle for a certain monetary offer on the grounds that, say, $150,000 is good enough, even if I expect that if I were to hold out for a better offer, I could probably get a larger amount. On the other hand, were a buyer to write down two offers, one of $150,000 and one of $160,000, it would be obviously irrational, *ceteris paribus*, to take the lower offer. The advocate of GSR cannot explain this difference; after all, if what made the first offer rationally permissible was simply the fact that it was good enough, it

---

[7] I am assuming that the choices are exclusive. In most situations, you would probably have a chance to look at both paintings.

should also make my choice of the lower offer permissible in the second scenario. The advocate of RSR has no problem, at least in principle, explaining the difference. The context that makes satisficing rational was present in the first case, but absent in the second. More specifically, in the first case, holding out for a better offer involves prolonged anxiety, unsettledness, etc.

However, once we made the advantages of RSR clear, we also made it apparently easy for maximizers to accounts for the phenomena. The maximizer can say that the agent is satisficing only from a *local* perspective: that she is, in our example, settling for a sale price that is merely good enough *relative to how much she could get for the house*. But getting as much money as she can for the house is only one of her goals and reflects only one aspect of her preference ordering. She also wants peace of mind, to avoid wasting her limited resources, etc. She is a local satisficer, but a *global* maximizer. In other words, she is settling for a good enough price since this way she maximizes her expected utility by taking into account also her preferences for not wasting too much time in the sale of the house, reducing anxiety, etc. The maximizer will gladly make a concession here. It might be that adopting a strategy of satisficing some goods is the best deliberative strategy. Given our limited resources, calculating expected utilities is prohibitively expensive; a global maximizer should often use local satisficing as a decision procedure.[8]

At first, this response seems to miss the mark. After all, the intuitive claim is that someone is permitted to choose from a large range of options that are good enough but not required. The idea is not that an agent *must* sell her house once she reaches a certain price, while knowing that she could get a better price. She *may* do so, but she is not required to do it. The intuitive thought is that it would be no less rational if she held out and waited to get a bit more money, and this would be as rational as if she held out for even more money, etc. Maximization would require either an implausible indifference curve among these options or that only one such action (or at least many fewer actions) would count as rationally acceptable.

---

[8] For this strategy, see D. Schmidtz, "Satisficing as a Humanly Rational Strategy," in Michael Byron (ed.), *Satisficing and Maximizing: Moral Theorists on Practical Reason* (New York: Cambridge University Press, 2004); H. A. Simon, "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* 69 (1) (1955), 99–118; and H. A. Simon, *Administrative Behavior*, 4th edn (New York: Free Press, 1997). Schmidtz and Simon are not defending what I will be calling "proper satisficing," so this is not a criticism of their views in particular.

But the maximizer can also reply to this objection. First, the need for a heuristic itself can already justify some kind of leeway in the range of options that will count as rationally optimal. If using a certain kind of heuristic is optimal, and the heuristics do not distinguish between various options that would otherwise have different utilities, then it is rational to choose any of the options that would be recommended by these heuristics. Moreover, arguably our ranking of states of affairs is incomplete or imprecise. The fact that none of these stopping points is rationally required might be just a reflection that our attitudes do not fully determine a precise tradeoff between, say, profit and convenience.

We can now see that there are at least three different rationales for satisficing that are fully compatible with a more general maximizing theory of rationality. First, *local* satisficing in relation to a certain good (trying to get enough money, bake a good enough cake, etc.) is compatible with *global* maximizing when all the agent's preferences are in. Second, a maximizing general theory can allow that we have *limited resources*, and thus satisficing might be an optimal decision procedure; that is, the decision procedure that has the highest expected utility. Finally, satisficing might be allowed in cases within "gaps" and "imprecisions" within our preference ordering.[9] Given the high level of idealization involved in the assumptions of orthodox expected utility theories, it is hardly contentious that they are not suitable to be universal decision procedures for actual agents. Thus, it is relatively easy to agree that, at least in some contexts, satisficing rules are our best guides. Given human agents' resource limitations and imperfections, an agent who adopts satisficing rules might do better, even if "doing better" is understood as "maximizing expected utility." We can assume widespread agreement with the claim that it makes sense for agents in certain circumstances at least to be *guided* by a rule of satisficing. However, insofar as the correctness of a satisficing rule depends on the fact that following such a rule maximizes expected utility, it is not, in my terms, a "proper" satisficing rule. A "proper" satisficing rule is a rule of instrumental rationality whose validity does not depend on its maximizing potential. From here on, I examine the possibility of proper satisficing rules.

---

[9] Strictly speaking, orthodox rational choice theory requires that one's preference ordering be complete. However, many of those who are otherwise sympathetic with maximizing theories of rationality have expressed doubts about the validity of this requirement. See, for instance, J. M. Joyce, *The Foundations of Causal Decision Theory* (New York: Cambridge University Press, 1999), and M. Kaplan, *Decision Theory as Philosophy* (New York: Cambridge University Press, 1998).

When to satisfice

A difficulty for satisficing arises from the above discussion. Suppose I identify, in a choice situation, an outcome O_s that is "good enough" and another one, O_m, with greater utility or value, that is the optimal point. Given that I know that O_m is better than O_s, how could it be rational to choose O_s over O_m? Here it seems that saying that O_s is good enough is completely unsatisfactory. Rather than insisting that rationality requires only that we choose an outcome that is good enough, most advocates of satisficing try to identify features of certain choice situations that are not given their due by the maximization procedure. So perhaps there is a virtue of moderation that requires us not to maximize;[10] perhaps our psychology is such that we evaluate things from perspectives not fully accounted for by maximizing;[11] perhaps satisficing (or at least some version of it) might make it easier to accommodate value incommensurability.[12] These views have a common strategy: they accept that, from a certain perspective, maximizing provides the correct choice of action, but given peculiar features of human psychology or the nature of value, the maximizing conception turns out to be a partial one. A maximizing conception, on these views, makes certain important considerations invisible to the rational agent. However, these approaches have some shortcomings: they depend on substantive assumptions about the nature of value or about our "evaluative psychology," or they need to defend satisficing as a second-best strategy for beings with limited cognitive powers. I propose here a different view. It is part of the very nature of the instrumental theory of rationality that a maximizing conception cannot be a general theory of instrumental rationality, especially if we focus on the rationality of *actions* rather than of *decisions*. We need only very minimal assumptions about ends to show that maximizing is an option only for a subset of actions and that a more general theory of rationality must incorporate at least certain permissions that are incompatible with a general maximizing theory. I hope to show that a rational agent could not obey the principle of instrumental reasoning in most circumstances without being guided by a rule of satisficing.

---

[10] M. Slote, "Satisficing Consequentialism," *Proceedings of the Aristotelian Society,* supplementary vol. 58 (1989), 139–63. Slote has changed his views on this matter since: see Slote, "Two Views of Satisficing," in Byron, *Satisficing and Maximizing*, 14–29.

[11] M. Weber, "A New Defense of Satisficing," in Byron, *Satisficing and Maximizing*, 77–105.

[12] M. Byron, "Simon's Revenge: or, Incommensurability and Satisficing," *Analysis* 65 (288) (2005), 311–15.

Orthodox rational choice theory requires that our preference ordering must be fully determinate so that for any lotteries L_1 and L_2, either $L_1 \geq L_2$ or $L_2 \geq L_1$. As we said above, this is an assumption that many philosophers find suspect. It is often agreed that a rational agent need not have a definite preference between some outcomes, let alone between probability distributions over outcomes. There is widespread agreement that the agent's attitudes such as preferences or ends might be fully rational without being fully determinate. More importantly, orthodox rational choice theory idealizes away an important set of ends and projects we have: namely, indeterminate (or not fully determinate) ends that are pursued not in a moment but through a long series of momentary actions.[13]

All, or at least nearly all, our ends are indeterminate and need to be pursued through actions that extend through time. Even rather simple actions such as baking a cake cannot be completed in a single momentary action. Given that I have no access to a miracle mixture, I must bake a cake through a series of actions that will span intervals in which I will be doing nothing that is instrumental to baking a cake (taking a break, answering the doorbell, etc.).[14] Moreover, our end in baking a cake will be in various ways indeterminate or vague: what counts as an edible cake (or a cake at all), how long it should take to bake, how large it should be, etc., is not precisely determined.

Suppose, for instance, that Mary has the end of running a marathon. There is a great deal that is indeterminate about her end. First, of course, there are many opportunities to run a marathon. Doubtless, if she waits too long to do it she will not achieve her end; however, there is a range of time in which any marathon she completes within that time would count as reaching her goal. She also does not necessarily set on a particular time she thinks she needs to complete a marathon, even if there are obvious

---

[13] I discuss these ends in S. Tenenbaum and D. Raffman, "Vague Projects and the Puzzle of the Self-Torturer," *Ethics* 123 (1) (2012), 86–112, and S. Tenenbaum, "Akrasia and Irrationality," in T. O'Connor and C. Sandis (eds.), *A Companion to the Philosophy of Action* (Malden, MA: Wiley-Blackwell, 2010), 74–82. They are often referred to in these papers as "vague ends" or "vague projects." My view is that vague ends constitute a proper subset of indeterminate ends, but nothing here depends on this. The discussion below is a summary of the more detailed discussion in these papers.

[14] On this issue, see K. Falvey, "Knowledge in Intention," *Philosophical Studies* 99 (1) (2000), 21–44. Tenenbaum, "Akrasia and Irrationality," and S. Tenenbaum, *Appearances of the Good: An Essay on the Nature of Practical Reason* (Cambridge University Press, 2007).

scenarios in which she would not have achieved her goal of running a marathon simply because it took her too long to cover 26.2 miles. Moreover, in such cases, how to achieve the goal is also vague and indeterminate. It is not clear how many minutes of training Mary needs to be ready to run a marathon, and when her training will happen; she needs a balanced diet, but she can sometimes eat cake and ice-cream; as she is running the marathon, she might start very fast, or somewhat fast; etc. In fact, we might notice something important about Mary's momentary decisions as she is pursuing this end. Suppose she needs to decide whether she is going to go out right now for a run. Given the indeterminacy and vagueness of the situation, it seems that at any time she might, say, spend one more second singing instead of going out for her run.

Of course, if she does this continuously, at some time it will be too late to go for a run on this day, but since one missed day of running is unlikely to make a difference to her chances of being ready to run a marathon, these choices will not be incompatible with successfully running a marathon. Surely, if she does this every day, at the end she will no longer be in position to run a marathon in a way that would count as successfully fulfilling her goal. But there might be no precise point at which this happens, or if there is, Mary might never be in a position to know that she is at that point. And, of course, if she is past the point at which she can still get herself ready to run a marathon, it would be perfectly rational for her to keep on singing; after all, since she has already failed in her end of running a marathon, there is no point in missing out on an opportunity to sing. In such a case, the following might obtain: given that postponing running for a second does not make a (significant) difference to her aim of being ready to run a marathon, and given that, say, at any moment **m** she prefers singing at **m** over running at **m**, it might be that for a certain interval, at each moment it is rational for her to sing rather than run. However, if she chooses to sing at every moment during that interval, she will not be ready to run a marathon, and she might prefer to be able to run a marathon than to sing on any number of occasions, and this preference might be one that she has stably throughout this whole interval. Assuming that there is nothing wrong with Mary's ends and preferences,[15] if Mary always chooses to sing, her

---

[15] One might think that we have already indicated what is wrong with Mary's ends and preferences: they are vague and indeterminate. However, given that nearly all our ends are indeterminate in this way, one needs to give a compelling reason why indeterminacy or vagueness suffices to make an end irrational.

choices will be overall irrational without any of her momentary choices being irrational. After all if she never trains for a marathon her choices lead her to an outcome that she regards as clearly unacceptable, given her original options.[16]

Given these facts about Mary's ends, we can say that her ends allow for the possibility of *top-down irrationality*;[17] namely the possibility that the rationality of a set of momentary actions does not supervene on the rationality of each momentary action. But since Mary's end does not *require* that she runs at any particular moment, but if she fails to do this at every moment she will not have pursued efficient means to her end, it seems that what instrumental rationality demands with respect to such ends is that she be permitted to run even if running is not her most preferred momentary action, since the cumulative effect of always choosing her most preferred action leads to the choice of a less preferred option. I propose that actions that one performs within a range of time when one is pursuing a long-term, indeterminate end or project need to be evaluated from two distinct (but compatible) perspectives; a perfectly (instrumentally) rational agent is never deemed irrational by either perspective. A punctate perspective evaluates whether a momentary choice or action is rational, given the agent's various preferences and ends. For our purposes, all that matters is that a punctate perspective will permit the agent to choose her most preferred momentary action, but will also permit the agent to forego her most preferred action in favor of pursuing actions that are constitutive of the pursuit of the agent's long-term, indeterminate ends. Were it not for such permissions, the agent in the above example would need to sing at every moment, and thus would not be able to run a marathon (an option that, according to our stipulation, the agent considers to be unacceptable at every moment in the relevant time interval).

The *extended* perspective evaluates whether enough of these permissions have been exercised and thus the end has been achieved. So if our

---

[16] One must tread carefully here, given that, if I am correct about Mary's ends and preferences, her preferences are not transitive. I think the right thing to say here is that Mary is irrational if she never trains for a marathon because she ends up with a choice that she knows to be *unacceptable*. For the notion of acceptability in question, see Tenenbaum and Raffman, "Vague Projects."

[17] I discuss the possibility of top-down irrationality in S. Tenenbaum, "The Vice of Procrastination," in Chrisoula Andreou and Mark White (eds.), *The Thief of Time* (New York: Oxford University Press, 2010), 130–50; Tenenbaum, "Akrasia and Irrationality"; and Tenenbaum and Raffman, "Vague Projects." The points I make in the rest of this paragraph are discussed in more detail in these works.

agent does not train on enough occasions, she will be unable to run the marathon. Although at each point she will have chosen something that from the punctate perspective she was permitted to choose, she will have violated a requirement from the extended perspective: the requirement to take the means to an end that she held stable through the whole extent of the time interval in question. So although we cannot locate her irrationality at any particular moment in time, the extended perspective deems that she has acted irrationally by failing to take the necessary means to one of her long-term ends. This is obviously just a very rough sketch of how principles of instrumental reasoning apply to long-term, indeterminate ends, but hopefully it will be enough for our purposes.

Getting some satisficing

Nothing we have said so far rules out the possibility that Mary could also train for the marathon non-stop and never spend any time singing. The punctate perspective allows an agent to exercise permissions not to choose her most preferred action. But since we imposed no limits on how often those permissions can be exercised, Mary could *always* choose to go running rather than singing. Given what we know of human nature, this is rather unlikely, but nothing we have said so far would make it irrational for Mary to choose this alternative. But assuming that Mary does have singing as one of her ends, always exercising such permissions would also be a violation of instrumental rationality.

Let us look for a moment at what seems to be an intrinsic limitation of the principle of instrumental reasoning, at least if we think of that principle as the sole principle of instrumental rationality. Mary has two ends: singing and running a marathon (or the instrumental end of getting ready to run the marathon). At this level of abstraction, these ends are fully compatible. But each of these ends has a certain internal structure: for each of them there are better and worse realizations of the end. As far as the end of running a marathon is concerned, a faster marathon is better than a slower one, certain courses are better than others, etc. If no other end were relevant to one's life, one would train as much as one could without risking injury, would eat only those foods that would not interfere with one's training, would never wear uncomfortable shoes, etc. One's life purpose in this case would finish as one crossed the finish line.

We can think of the somewhat different end of being a marathon runner, in which this gloomy outcome would not be necessary. The end

of a run would only signal the beginning of a new training regimen in preparation for the next one.

Like being a marathon runner, singing has an internal structure that never fails to give meaning to one's life. There is no point at which singing reaches its natural stopping point, but it is always "complete" at any point when one stops singing. If I stop my marathon training too early, I will have failed to achieve the end for which I was training (namely, completing a marathon). But no matter when I stop singing, I will not have failed to realize the end of singing. And if I keep training for a marathon after I have run it, I will just be wasting time (unless I aim to run another one); but if I keep singing, I will keep realizing the end of singing. "Singing" is one of the verbs that fall under the category that Vendler calls "activity terms": roughly, those terms that refer to activities which are complete in themselves and do not have an end outside themselves.[18]

But it now seems that circumstances will abound in which the principle of instrumental reasoning will fail to provide any determinate guidance to an agent. After all, if the agent has conflicting ends, she will have to violate the principle of instrumental reasoning. On this understanding of the end of singing, there is nothing about it that dictates that its pursuit should ever terminate, and thus the pursuit of any end that cannot coexist in time with singing conflicts with my unrestricted end of singing.[19] Moreover, as we know all too well, singing can be done better or worse. Admittedly some shower soloists seem completely oblivious to the quality of their singing, but, in general, having singing as an end, and caring about how it sounds, go hand in hand. If you are among those unfortunate souls who cannot run and sing at the same time, or if you cannot do both just as well concomitantly as in isolation, you cannot, insofar as you are rational, adopt the (unrestricted) end of singing and the end of being a marathon runner. However, the principle of instrumental reasoning says nothing about how you should proceed to revise your ends in such cases.

---

[18] See Z. Vendler, "Verbs and Times," *Philosophical Review* 66 (2) (1957), 143–60, and see also A. P. Mourelatos, "Events, Processes, and States," *Linguistics and Philosophy* 2 (3) (1978), 415–34. My usage departs slightly from Vendler's. One could take issue with many of the things I say about "singing" and "training for a marathon," but (hopefully) these issues will not matter to the general argument.

[19] One could have as an end singing just for a minute, or from time to time. My only point is that nothing about having singing as an end mandates this structure. However, it is worth noting that for the main purpose of the chapter such ends could also be used; we would only need to fill in more of the story to ensure that there was potential conflict.

Such cases of conflict explain the attraction of moving from a traditional conception of instrumental rationality to a more modern one.[20] The traditional conception of rationality takes the instrumental principle of rationality to be the sole principle of rationality, so it has willing (or pursuing) ends, rather than preferences, as the fundamental attitude to which the principles of instrumental rationality apply. In its primary application, it is not comparative; it simply determines that certain means must be taken (or suffice) for the achievement of a certain end. In this primary instance the only comparative judgments that follow from the principles of instrumental rationality are rather trivial ones; they (nearly) all have the form "it is better to pursue necessary means X to your end Y than not to pursue it." If our ends do have the kind of internal structure I have been suggesting above, they are capable of generating some preferences, and instrumental rationality will be able to make some non-trivial comparative judgments. Typically, when one is building a house as an end, one aims to build an excellent house. So, arguably, commitment to the end already commits one to prefer means that deliver a comfortable house over means that deliver a barely liveable one. The nature of my end in building a house and the instrumental principle suffice to determine a certain partial ordering internal to this pursuit. For instance, *ceteris paribus*, given the empirical facts about construction, insofar as a rational agent aims to build a house, she will prefer to use bricks over papier mâché.

It is worth distinguishing between the claim I am making here and a similar claim made by Korsgaard.[21] Korsgaard says that because a house has constitutive standards, in adopting the end of building a house we are also committed to building a good house. Even a shoddy builder, if he is building a house at all, must be bound by the norms of good house

---

[20] Of course moving to a theory that favors preference ordering as opposed to ends provides for greater ease in accounting for risk. If what we say here is correct, this is an illusory advantage; if rationality does not mandate anything that approximates a preference ranking in the case in which we have certainty, it will certainly not do it once we introduce risk. However, the view put forward here still allows that an end could determine a preference ordering even with the sufficient precision necessary for a utility measure. Given that we allow that an end does generate a preference ordering internal to itself, there is no reason to think that there is in principle any limit to how precise or fine-grained such ordering would be. Similar things apply to the case of very general means such as money or power. Of course, one would like a more general theory of rationality under risk, but we have to leave this task for another occasion.

[21] C. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (New York: Oxford University Press, 2009).

building.[22] I agree with Korsgaard that there is a characteristic function of building a house that determines to some extent that some houses are better than others. Moreover, typically, when one adopts the end of building a house, one adopts the end of building something precisely because it has this function. In such cases, abstracting from competing ends, adopting such an end requires that one choose the better house over the worse house. However, nothing we say here commits us to the view that *everyone* who builds a house must adopt precisely this end. Someone might be building a historically accurate house or a highly profitable house; the fact that a better house is insulated with polyethylene ground cover might be of no practical significance to either builder. There is a broader question of whether any end we adopt will provide grounds for similar comparative judgments relative to the end; again, for our purposes, we need not commit to an answer to this question. But it is important to note that we should not expect that such comparative judgments would always yield anything approximating to a complete ordering: although some houses are clearly better than the others, there are many that are not clearly comparable with respect to the end of house building.

At least in the case of a single end, the comparative judgments needed are, at least in part, readily available. But in exactly these cases we do not need to appeal to an independent preference ranking; the instrumental principle and the ranking generated by the end will fully account for the relevant rational requirements. However, when we come back to different ends, it is not clear what can we avail ourselves of to determine a ranking of options. The two ends, singing and marathon running, conflict, but there is no end of singing-while-marathon-racing that generates an internal ordering (or if there were, we would be looking at an altogether different case).

Since I cannot be fully committed to singing and to running a marathon such that I always choose both the best option from the point of view of running and singing, instrumental rationality requires that I do not pursue both singing and running a marathon in this unrestricted manner. But we cannot appeal to the hyphenated end to determine what would be the sufficient or best means to pursue the end of singing-while-marathon-racing; there is nothing in the nature of either end (or both ends considered together) that determines how the conflict should be

---

[22] C. Korsgaard, "Acting for a Reason," in Korsgaard (ed.), *The Constitution of Agency* (New York: Oxford University Press 2008), 207–31, and Korsgaard, *Self-Constitution*, 113.

resolved. So, the only unambiguous recommendation in this case for a rational agent is to revise her ends, given that the pursuit of each end in an unrestricted manner is not a coherent possibility.

But nothing so far follows about how they should be revised: that is, how much importance should be accorded to each end such that I know when I should choose to sing and when I should choose to train; when I should rest and when I should take a risk of injury and navigate through treacherous terrain so that I can keep my appointment with my singing coach; etc. One might say that given what has been said so far, any way in which I revise my ends is fine as long as I no longer have conflicting ends. I have much sympathy for this view, but if this is the correct view, we have no reason to appeal to preference ordering. After all, once the ends are no longer in conflict, we can go back to using the traditional conception of instrumental rationality, and let our decisions be guided by the ends themselves. So, for instance, if I revise my ends so that I am just training to complete one marathon and to sing a few hours per week, it seems that I need no further rules of rationality than the instrumental rule: pursue the means to each of these ends. As we will see in a moment, things are a bit more complicated, but so far we have not seen why a preference ordering would be necessary.

Another route to introducing preference ordering would be to argue that (instrumental) practical reason cannot be neutral regarding adjudicating between potentially conflicting ends. In other words, there must be something in the nature of these two ends or the agent's attitudes towards them that determines their relative importance and consequently a preference ordering. There are two basic arguments one can give for such a conclusion. One can argue that there is some feature of the attitude of having an end that will necessarily give rise to a preference ordering. Such an argument might proceed from a general conception of the nature of motivational states, or desires. On this view, desires are not distinguished from each other simply by their content; they are also held with different degrees of intensity. My desire to sing might be very strong, while my desire to run a marathon might be very weak.

We cannot examine such views in detail, but the problems should be quite obvious. If "strong" and "weak" are being used in a non-question-begging way (namely, if we are not identifying "strong" with "there is a strong instrumental reason to pursue it"), it is unclear why instrumental rationality would recommend that preferences track strengths of desire. One could add a principle of rationality that mandates that one pursue one's desires in proportion to strength, or a theory of reasons which

implies a requirement of this form, but such principles and theories are not part of a theory of instrumental rationality.

Let us grant that insofar as it is possible to ascribe a complete preference ordering to an agent, then an instrumentally rational agent will maximize utility (when utility is understood as a measure of preference). Suppose, for instance, we can identify a preference ordering for an agent for choices between apples, pears, and bananas. We assume that in such a case the agent will be instrumentally rational insofar as she maximizes utility with respect to her choices among these items. Here a maximizing rule would determine the choices of the agent in a way that goes much beyond any constraints that can be imposed by the principle of instrumental reasoning. I have been suggesting so far that nothing about the nature of practical rationality implies that a rational agent, even an ideally instrumentally rational agent, must have a preference ordering that is determinate enough to imply any constraints beyond those determined by the principle of instrumental reasoning.

But one might now think that it is exceedingly easy to attribute a preference ordering to a rational agent. In fact, we have an effective procedure to determine the preference ordering of any agent to any arbitrary degree of precision: namely, Ramsey's procedure.[23] The procedure asks an agent to make choices between various outcomes in hypothetical choice situations. So a decision theorist with too much time on her hands could ask me a large number of questions about various tradeoffs I would be willing to make between lotteries involving various episodes of singing and various training episodes (and other things that I might be interested in) until she can form a complete preference ordering. But the in-principle availability of such a Ramsey procedure does little to establish the importance of its potential outputs for a theory of rationality. There is no guarantee that my answers are tracking anything significant: there is no reason to think that they are tracking any robust behavioral dispositions, let alone anything that has any relevance to our theory of instrumental rationality.[24]

One might argue that there must be something about the ends that determines the relative importance of each end for the agent, so that in having ends we must be committed to at least some kind of rough preference ordering. After all, we do not take all our ends to be equal.

---

[23] F. P. Ramsey, "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays* (New York: Harcourt, Brace, 1931), 156–98.
[24] See Joyce, *The Foundations of Causal Decision Theory*, for a similar point.

We certainly think that some of them should be given very high priority over others: my end of providing a good education for my children takes considerable precedence over my end of growing sunflowers in my garden. It is undeniable that certain ends are more important than others for the agent; the question is, however, whether this fact should supervene simply on the attitudes the agents have in holding each end. In other words, the proposals we are looking at claim that *simply in virtue of having each of these ends* (or at least of having these ends in a certain manner), the agent will also have preferences that express the relative importance of each of these ends to an arbitrary degree of precision. But this is rather implausible. The relative importance of each end should be understood as a further attitude that the agent might or might not have: namely, a second-order end to give priority to one end over another.[25]

A different line of argument in favor of the idea that a theory of instrumental rationality should rely on preference orderings rather than ends is more interesting and important for our purposes. If an agent's ends cannot be each pursued in an unrestricted manner and yet they do not determine a preference ordering, how would a rational agent determine what to do in such situations? Even if all that an agent is required to do is to revise her ends, she must determine how she will adjudicate between the incompatible demands of the original ends. The peace treaty between the warring ends must determine the borders as precisely as possible; it should determine which of my actions will belong to singing and which to training for the marathon. But this means that the revised ends are just a preference ordering determining in which situations I choose singing over training and vice versa.

I do not think that there is any determinate answer to the question of how a merely instrumentally rational agent must proceed in cases in which her ends conflict; in my book, instrumental rationality requires nothing more than that the agent revise her ends so as to end the conflict. The true *substantive* theory of rationality or the good might have something to say about this; perhaps singing is more important than running; perhaps we need to live a balanced life; possibly a human being must do whatever gives her most pleasure in such situations. But these considerations are beyond the scope of a theory of instrumental rationality.

---

[25] Of course, these attitudes could be grounded on other attitudes or judgments of the agent: perhaps they are grounded on the agent's evaluative belief or the agent's identification with some ends but not others. But this attitude is not rationally required simply by having (or by the manner one has) the ends in question.

However, I will not argue for these claims here. What matters for my purposes is that we can end the conflict without relying on anything as richly detailed as a complete preference ordering. One could also simply adopt instead restricted ends such as "enough singing given that I am also training for a marathon" and "enough training given that I am also a singer." These are vague or indeterminate ends, but, as we said above, there is nothing intrinsically irrational about having ends with this kind of structure. And they need not conflict; one could succeed in the pursuit of each of them. In fact, this is the most natural (and certainly the minimal) way of resolving the conflict: restricting each end in light of the pursuit of the other.

This minimal revision leaves us with two vague or indeterminate ends, and because of this structure there will be no point at which it will be the *maximum* amount of singing I could have done while still being able to train for the marathon. At each point, singing for a millisecond longer would not affect my marathon running, or, at the very least, I would never be in a position to know that it would. If at each millisecond, I had the thought "one millisecond longer will make no difference" and kept on singing, at some point I would realize that it was too late to train for a marathon. Given the possibility of top-down irrationality discussed above, the pursuit of our revised ends to a *maximum*, insofar as we can make sense of the idea, or to pursue our most preferred option at every moment, would be necessarily self-defeating. Thus when evaluating over a period of time whether I was successful in my pursuit of the ends of singing and training for a marathon, the criterion for success could not be whether I reached the greatest amount of singing or training or a maximum point in some function that combines both. I will be successful in my pursuits instead if I have done – well – *enough* singing and training. And since non-accidental success in the achievement of an end should suffice for determining that an agent does not run afoul of any principles of instrumental rationality, we can say that instrumental rationality requires no more than that I satisfice. That is, I should pursue singing until my singing is "good enough" and I should train until my training is good enough; rationality is here judged in terms of whether the agent has passed a certain vague minimal point, not in terms of whether she has reached a maximal point.

This might look like local satisficing, rather than proper satisficing; after all, we are not saying that the agent should choose an action that is good enough *simpliciter* or that has more than a certain minimum amount of utility, but rather that in pursuit of each separate end, she

should ensure that she reaches, as it were, a minimum level of each. However, it makes sense to say that a theory should recommend only local satisficing, if local satisficing can be contrasted with global maximizing. But there is no level in the theory in which any kind of maximization is taken to be an ideal.[26] If there is a global measurement of assessment it is also a satisficing one, at least in a natural reading of "satisficing." Global success will be something like "achieve all ends to an acceptable level" or "engage in enough singing, enough training, etc." In either case, global requirements will also be requirements to achieve a certain (vague) minimum threshold.

Notice that we came to the conclusion only by looking at the structure of the ends of a rational agent; we made no assumptions about her computational powers, how much leisure she has to crunch numbers, etc. In sum, a satisficing rule here is the right rule given a perfectly rational set of ends, or a perfectly rational structure of pursuits. If we are right that there is nothing amiss with such ends from the point of view of instrumental rationality, we have shown that satisficing is not just a blunt tool used by limited rational beings operating in a world full of friction. It is, at least in some cases, also the correct rule for a perfectly instrumentally rational agent acting in a frictionless world.

Getting more satisficing

So far we looked only at what we called above "the extended perspective." We saw that in order for an agent to realize ends that are potentially conflicting and indeterminate or vague, the agent will have to use a satisficing criterion of success. But when we look more closely at the structure mandated by ends that introduce the possibility of top-down

---

[26] Of course, this rules out the possibility that satisficing can be defined in terms of a threshold of utility. H. S. Richardson, "Satisficing: Not Good Enough," in Byron, *Satisficing and Maximizing*, raises this as a more general problem for satisficing. I have tried to avoid relying on a specific satisficing rule, but I should note that, obviously, the views presented here are not compatible with defining satisficing in terms of any kind of precise threshold. To make it fully compatible with the views defended in Tenenbaum, "The Vice of Procrastination," and Tenenbaum and Raffman, "Vague Projects," we should define the basic rule of satisficing relative to an end E something like "Choose an option that is not an unacceptable realization of E and does not prevent the acceptable realization of your other ends." Further refinements are needed, but here I just want to argue for the general plausibility of defending a satisficing rule in this manner. See Tenenbaum, "The Vice of Procrastination," and Tenenbaum and Raffman, "Vague Projects," for further discussion of the notion of "acceptable" employed here.

irrationality, it becomes clear that in the punctate perspective, something like a satisficing standard also plays an essential role in determining what should count as rational action. Let us take, for instance, decisions I make in various circumstances in which I would prefer to do something on each occasion, but for which the cumulative effect of always choosing the most preferred (momentary) option is disastrous. Let us take, for instance, eating. I do not (or at least let us assume I do not) have a long-term vague end or project with respect to gastronomic pleasures. However, at various moments I enjoy eating, especially fatty or sweet foods. If we look at my momentary decisions, I prefer, over a long period of time, to continue eating these foods rather than stop. After all, one more bite of food will not make a difference to my overall health and looks, and it takes quite a bit of time for me to be so sated that I prefer to refrain from having just one more bite of something totally delicious.

Yet this pattern of activity, as we know only too well, has disastrous consequences.[27] If I am right about the possibility of top-down irrationality, this is clearly a pattern of choice situations that allows for this kind of irrationality. At every momentary choice, it might be rational to choose to eat more, but if I always make those choices, my health will certainly suffer in ways that are not compensated by the additional gastronomic pleasures. A rational agent will stop maximizing at some point, even though at this point he could do better (by taking just one more bite). In the view I proposed above, the pursuit of the end of leading a healthy life will generate permissions in the punctate perspective to pursue this end even when it does not maximize utility. In fact, it seems that an agent could exercise this permission whenever confronted with some kind of fatty or sweet food: the agent might simply refuse anything that is not a constitutive item of a healthy diet.

But this would be a sad aspect of one's life, and nothing about the ends of the agent dictates such sadness. The end of having a healthy life, for most of us, is fully compatible with a few indulgences. However, pursuing one's most preferred option (or the best option) at every single momentary choice *is* incompatible with a healthy life. The reasonable agent will indulge here and there while making sure that his indulgences are not unduly jeopardizing his health. But the only way to do this is to stop indulging when one has done it *enough* times. In other words, a rational agent does not exercise a permission to forego a gastronomic pleasure

---

[27] This pattern is discussed in more detail in Tenenbaum and Raffman, "Vague Projects."

from here on (by, say, letting the server take his plate) until he has enjoyed *enough* gastronomic pleasure, or until he has failed to exercise this permission enough times. Although it is harder to characterize this rational constraint on more traditional satisficing terms, something like the following satisficing rule seems correct:

> In the relevant context, exercise permissions to pursue end E only to the extent that doing so allows you to pursue often enough your momentary preferences.

If I am right, "often enough" cannot be replaced by a precise threshold. But if I am right, postulating precise thresholds is the wrong way to characterize and defend the intuitive idea that, even when better options are available, the rational agent is often satisfied with an outcome or action that is good enough.

Conclusion

If we start with the assumption that expected utility theory is the overall correct theory of rationality, a decided improvement over earlier theories of instrumental rationality, satisficing seems like either a quaint request to be contented with less than we could have, something reminiscent of monkish ideals of self-mortification and abnegation, or at least a resigned recognition of our limited cognitive powers. I have been urging that the mistake is in accepting this assumption without further ado. Pursuing what is just good enough (or some precisification of this idea) is the only possible advice in the context of the extended pursuit of indeterminate ends. When we think of the rationality of *actions* rather than of *decisions* or *preferences*, we wonder how often anything is done that falls outside this context. After all, actions are extended through time, and our actions are particulars, whereas the content of our intentions is general. Perhaps long-term indeterminate ends are not only something that is possible for a rational agent, but are just the kind of ends that rational agents (or at least rational agents that do not have intellectual intuitions) necessarily have. But defending this final claim is a project for another occasion.