4

Intransitive Preferences, Vagueness, and the Structure of Procrastination

Duncan MacIntosh

Procrastinating is irrationally failing to do something in good time. According to Chrisoula Andreou, agents are sometimes induced to procrastinate by having intransitive preferences, possibly in combination with vagueness in the circumstances of choice.[1] I argue that her model cannot explain procrastination and that its true explanation is in things already familiar from the literature on weakness of will.

PROCRASTINATION AS IMPRUDENT DELAY

One is guilty of procrastination when one has a self-acknowledged best reason in favor of doing something today but instead does it tomorrow (or never). This should not be confused with excused delay, where, for instance, one physically could not do the thing until tomorrow, nonculpably did not realize that it would be better done today,[2] had not figured out how to do it today, had ambivalence about whether one really wanted it done, or was prevented from doing it by some irresistible urge outside one's desire structure, that is, by something that prevented one from being fully an agent. Another species of excused delay is where one ceased to have a best reason to do the thing in question because one underwent a rationally faultless change in what one desired; for good reasons, perhaps from further experience, new second-hand factual knowledge, reflection on or deduction from one's information, or changed circumstances, one changed one's mind about one's goals or had a rationally permissible

1. Andreou's theory has unfolded over a series of papers: "Instrumentally Rational Myopic Planning," "Going from Bad (or Not So Bad) to Worse," "Environmental Damage and the Puzzle of the Self-Torturer," "Temptation and Deliberation," "Understanding Procrastination," "Environmental Preservation and Second-Order Procrastination," and "Making a Clean Break: Addiction and Ulysses Contracts."

2. By nonculpability, I mean that one was not guilty of self-deception, took suitable care in investigating the facts, and so on.

OUP • UNCORRECTED PROOF

change of heart and so no longer had a practical duty to take the best means to the end at which one had been aiming.

Nor should procrastination be confused with prudent delay, where one puts off doing until tomorrow what one could have done today because it would be better done tomorrow (e.g., one would be able to do a better job of it) or because one had other ends that one ranked cumulatively at least as highly and that could best be attained by putting off the attaining of the end in question. (In fact, delay resulting from a rationally permissible change of heart about one's goals might better be classed as prudent delay rather than excused delay.) Procrastination proper is imprudent delay, where one puts off until tomorrow what one admits would, everything considered, be better done today. If one's reasons are grounded in one's desires and beliefs, then one acknowledges that, given these, doing this thing today is the best means toward servicing one's overall attitudinal structure. This is the only genuine form of procrastination: procrastination is necessarily irrational.

It is natural to class procrastination in with weakness of will, something that, since Donald Davidson, is often analyzed as acting to bring about an end one desires less than one's ostensibly most strongly desired end.[3] And it is commonly thought to be caused by such things as fear, nervousness or loss of nerve, boredom, aversion to the means needed to attain the end, wishful or unclear thinking, culpable ignorance or failure of foresight, failure to trace the logical consequences of one's beliefs, distraction or forgetfulness, lack of gumption, exhaustion, laziness, depression, and accidie or loss of affect. Some of these phenomena may need recategorization, depending on their analysis. For example, failure to do something to bring about a self-ascribed end because of aversion to the means may entail that one's overall preference ranking does not make the end whose attainment one is nominally procrastinating against, one's highest end (for, really, one's highest end was to not engage in that means), and so one is really engaging in prudent delay, not procrastination or any other form of weakness of will. Or maybe here we can speak only of procrastination relative to a nominal goal taken in isolation, not irrational delay relative to one's all-in ranking. Something similar might be said when someone fails to take the required means to her supposed ends from laziness: she may really prefer lying around. Meanwhile, failure to act in a timely way because of exhaustion may really be excused delay, as when exhaustion makes action impossible. And ostensible procrastination resulting from cognitive failures, such as forgetfulness, distraction, or not seeing the logical consequences of one's beliefs, may at least sometimes count as excused delay (one did not know there was something it would on balance advantage one to be doing) or as prudent delay (given how the facts seemed, delaying was rational).

3. For a survey of accounts of weakness of will, see Stroud, "Weakness of Will." As Jon Elster points out, this classification, in turn, requires explanation; see Elster, "Weakness of Will."

OUP • UNCORRECTED PROOF

ANDREOU'S THEORY: PROCRASTINATION AS INDUCED BY INTRANSITIVE PREFERENCES

Let us turn now to Andreou's work on procrastination. Note first that an agent's preferences are transitive just in cases where if she prefers *x* to *y* and *y* to *z*, then she also prefers *x* to *z*; otherwise, her preferences are intransitive. Andreou thinks that some cases of procrastination result from the agent's having intransitive preferences. Suppose you have some large goal, such as quitting smoking, losing weight, saving the environment, or writing a book. But suppose also that you find pleasant each individual act of smoking, eating, polluting, or lazing around not writing, and you believe that no individual such act will put your larger goals out of reach. Then, for each occasion on which you could indulge in smoking, for instance, you may prefer indulging then to beginning to quit smoking, even though, given the choice, you would prefer never indulging to always indulging, since always indulging would doom your larger goals—call this latter preference your *global preference*. You will then be led by pair-wise choices to having always indulged, with disastrous results—cancer, obesity, an unlivable environment, no book—results you will regret and yet be unable to reverse. You will be led to this even though, given the choice, you would take back all of the indulgences, because you prefer having never indulged to having always indulged (hence the intransitivity in your preferences). Your choices are then arguably irrational, because they result in your getting the opposite of what you globally prefer. Indeed, you may foresee this, even though, in each choice between indulging and working toward your larger goals, you take yourself to have justification for delaying the cessation of indulgences—for procrastinating.

Andreou takes her theory to be nicely illustrated by, and indeed based on, Warren Quinn's puzzle of the self-torturer.[4] In Quinn's story, you are connected to a machine that delivers electric shocks. It has 1,000 settings from lower to higher shock levels, with the levels becoming unbearable at some point in the progression of settings. Each week, you get to experiment with the shock levels, finding out which ones are intolerable. Then you have the option of either staying at the shock level you are presently at or moving up one level; you are never allowed to go back down. Each time you move up, you will get $10,000. Adjacent shock settings are stipulated to be indistinguishable in pain level to you, though settings far enough apart are distinguishable.

Arguably, each week, you will reason, "The next setting will not feel any more painful than the current one, and I will get another ten grand," so each week, you move up. But you will eventually wind up at level 1,000, which will be unbearable, even given the compensation of the money you will have; and you will desperately wish to go back to level 0,

4. Quinn, "Puzzle of the Self-Torturer," 198.

OUP • UNCORRECTED PROOF

but you will not be allowed to do so. Supposedly here, your preferences are intransitive, because you prefer each higher setting to the one below but also the lowest to the highest. Procrastination here is supposedly your delaying the stopping of the raising of the settings.

Procrastination with regard to quitting smoking, losing weight, and so forth, is the repeated delaying of acting on these larger goals one more time, on the rationale that one more time will not make a significant difference to their attainability but is itself desired. In the self-torturer case, your global preference, your preference for never raising the shock level rather than always raising it, is never expressed in action; only your pairwise preferences, your local preferences, your preferences to raise, are expressed. One appeal of Andreou's model is that it supposedly shows how the agent is seemingly rationally tempted into making certain choices and yet will regret them, culminating in a net irrationality.

I now begin a series of criticisms of Andreou's position. The first few speak to the intelligibility of the self-torturer scenario and whether there is any irrationality in having, or in acting upon, intransitive preferences, an irrationality that is necessary if the preferences are to explain procrastination conceived as necessarily irrational. I will argue that we can see delaying behavior of the sort Andreou has in mind as irrational only if we presume that the agent's preferences are transitive rather than intransitive. Later, I will critique another element of Andreou's view: the idea that procrastination can be explained by an inherent vagueness in what would count as an appropriate point to stop indulging and start fulfilling the more ambitious goals that figure in one's global preferences.

THE INCOHERENCE OF THE SELF-TORTURER SCENARIO: INDISCERNIBILITY, THRESHOLDS FOR SUBJECTIVE STATES, AND TRANSITIVITY

Andreou's model is ingenious and, on the face of it, plausible. But I think there are several fatal difficulties with it. First, it is doubtful whether the self-torturer scenario as usually described is coherent; and so it is doubtful that it can be used to make the points Andreou wishes to make. It is stipulated of the scenario that setting 1,000 is intolerable; furthermore, it is implied that what it means for a setting to be considered intolerable by an agent is that it is painful enough to not be worth the money he would earn from it. This implies that the agent prefers more money to less money, less pain to more pain, and that his aversion curve to increasing pain is steeper than his affinity curve for increasing amounts of money. It is also stipulated that adjacent settings are indistinguishable in terms of pain, but settings far enough apart are distinguishable, that some lower settings are tolerable, and that the number of settings is finite.

But now there is a problem. Suppose that when the agent is experimenting with the settings, he sets the level to 1,000, finds this is intolerable,

OUP • UNCORRECTED PROOF

and backs down the settings. It follows from everything we have said that if he backs the settings down far enough, he finds a tolerable setting. But this, in addition to there being only a finite number of settings, entails that, compared with setting 1,000, in backing down, he will find a first tolerable setting.[5] Suppose this is setting 995; then setting 996 will be intolerable. But then settings 995 and 996 are distinguishable in comparison with setting 1,000; 996 is intolerable, 995 tolerable. But then at least one pair of adjacent settings, 996 and 995, are distinguishable, contrary to the stipulation that adjacent settings are indistinguishable.

But the scenario remains of interest if we relax that stipulation, so we will; let us regard adjacent settings as distinguishable at the threshold of intolerability. There are now two possibilities: either the agent would stop raising settings each week at 995, or he would raise the setting to 996. If he stops and is rational in stopping, then his preferences are transitive: he prefers 995 to all other settings because it is the setting that best balances pain and financial reward. Higher settings are dispreferred because they are too painful, and lower settings are dispreferred because they involve less money. On the other hand, if he increases to 996, either he is rational in doing so, or he is not. If he is not, then, barring an excusing condition, he has procrastinated, irrationally delaying the ceasing of the raising of the settings. But then, contrary to what Andreou says, we would have no explanation for this in the intransitivity of his preferences, since we have just decided that his preferences are transitive; and we would have no explanation in the form of the vagueness of the intolerability of the settings, since we have already established that this cannot be vague. So, the agent has just made some mistake; perhaps he has been weak-willed in some standard sense.

Now, let us consider the other possibility: the agent increases to 996, keeps going up to 1,000, and then wishes he could go back to setting 0, having been moved in all of this by his preferences. It follows, as Andreou induces from his behavior, that his preferences are intransitive: he prefers each higher setting to the one just beneath it and at the same time prefers the lowest setting to the highest setting, and this preference intransitivity explains his behavior. But if his preferences really are intransitive, then, since his choices express his preferences, it is not clear that his choices are irrational, as they would have to be in order for his proceeding to setting 1,000 to count as procrastination, which is necessarily irrational. Whether he is being irrational depends on whether having, or choosing from, intransitive preferences is irrational.

WHERE IS THE IRRATIONALITY IN HAVING AND CHOOSING FROM INTRANSITIVE PREFERENCES?

If, as Andreou thinks, it is the intransitivity of the agent's preferences that explains his delaying something, and if this delay is to be seen as irrational—as

5. Thanks to Darren Abramson for discussion on this point.

OUP • UNCORRECTED PROOF

it must be in order to count as procrastination—then there must be something irrational about having intransitive preferences or about the choices they would induce their holder to make. I will consider later whether there is anything irrational about having intransitive preferences. But now I will argue that even if an agent's preferences are intransitive, there will be no obvious irrationality in the choices they will induce the agent to make; and if, as Andreou allows, to be rational is to best advance one's preferences, the agent's making these choice would seem indeed to be rational.

Consider an agent whose preferences look like this: A > B > C > A (where > signifies "is preferred to"). Imagine that he is never asked to choose among A, B, and C all at once (all-wise) but only two at a time (pair-wise). Then each choice justified by his intransitive preference structure seems perfectly rational, for each choice always improves his position relative to his current position. True, each choice also always leads him to an outcome he disprefers in comparison with the possible outcome of a further choice, and so he would always make yet another choice given the chance, and so on. If he has A, he would rather trade it for C; if he has C, he would rather trade it for B; if he has B, he would rather trade it for A; and if he has A, he would rather trade it for C; and so on. But this is not in itself even prima facie irrational.

If the self-torturer really has intransitive preferences, he, too, would be irrational to refuse to make the next advance. (Ironically, here, it would only be his refusing to advance that would count as procrastinating, that is, as his delaying a choice that would advance his preferences. And, contrary to what Andreou says, we would need recourse to something other than his preferences to explain his refusal—perhaps one of the standard explanations of weakness of will.) He would be irrational in refusing to raise his level because, at each setting, he prefers not stopping. So, he rationally should proceed to the maximum level. But of course, since his preferences are intransitive, he also prefers the minimum level to the maximum; and since he is not allowed to go back, he is stuck at the maximum level that he disprefers to the minimum. Has he not, then, been irrational?

No, there is no irrationality in his having gotten himself into a situation where he faces an obstacle to further choice, for there is nothing irrational about his wanting something he cannot have (or if there is, this is not the basis of Andreou's argument). Imagine that our agent's preferences are of the form A > B > C > A and that when offered a choice between C and B, he chooses B, and then when offered a choice between B and A, he chooses A. But then he is prevented from choosing between A and C, where, given the option, he would choose C. His preferences, plus his circumstances, have left him stuck with A when he would rather have C. But there is no evident irrationality in his finding himself in this predicament.

However, it might be argued that he finds some options acceptable and others unacceptable and that he ends up with an option in the latter

OUP • UNCORRECTED PROOF

category, though it was in his power to end up with one in the former. Surely, he is being irrational?[6] Again, no; for no matter what state he winds up in, he will find it unacceptable relative to another possible state. That is, no matter what state he is in, he would rather be in a different state. So, no state for him is such that it was acceptable while other options to which he moved were unacceptable. Therefore, no state is such that his failing to preserve it amounts to his being irrational for failing to preserve an acceptable state.

Nor does any of this change if, in order to respect the idea that the agent in Andreou's case is supposed to have been irrational for pushing on to a state he globally disprefers, we define one of the states in the intransitive preference cycle as the state of not having indulged at all and each of the others as being cumulative indulgences, so that the agent's preferences look like this (where < signifies "is less preferred than"):

$$(C) < (B \,\&\, C) < (A \,\&\, B \,\&\, C) < -(A \,\&\, B \,\&\, C) < (C)$$

Suppose this is the smoking case. The agent prefers having smoked a second cigarette, (B & C), to stopping smoking after the first cigarette, (C); he prefers having smoked a third cigarette, (A & B & C), to stopping after the second, (B & C); he prefers having smoked none of the cigarettes, –(A & B & C), to having smoked all of them, (A & B & C); and he prefers having smoked one cigarette, (C), to having smoked none, –(A & B & C); and so on.

He will always prefer each possible next indulgence to stopping; call these preference pairs his local preferences. And he will always prefer abstaining altogether to engaging in all of the indulgences; this is his global preference. Finally, he will always prefer an indulgence to refraining from all indulgences. Note that the mere fact that some of these preferences are global, since they involve all of the states over which his other preferences range, does not make their objects automatically highest-ranked states—in an intransitive ranking, there is no highest-ranked state. Anyway, there is nothing obviously irrational about this ranking or about making the pair-wise choices the agent would make were he presented with the options two at a time.

Meanwhile, if we introduce an obstacle at the point where the agent would elect to retract all of his indulgences, he cannot get what he wants; and since this is his global preference, he cannot get it satisfied. But again, there is no irrationality in that, for there is nothing here that privileges the global pair-wise preference over any of the other pair-wise preferences, the local ones in particular. Indeed, the agent would also have been unhappy had he been forced to stop after the first cigarette.

This means that there is, again, a reply to the worry that the agent finds some options acceptable and yet voluntarily moves from them into an

6. Thanks to Andreou for this objection.

OUP • UNCORRECTED PROOF

unacceptable option.[7] Again, no matter what state he winds up in, he will find it unacceptable relative to another state he could have had. So, again, no state is such that his failing to preserve it amounts to his being irrational for failing to preserve an acceptable state.

An agent's not getting his global preference satisfied is no worse for him than his not getting any of the others satisfied; it is not as if he had a transitive preference hierarchy with the object of this global preference ranked as the most-preferred object. There is no special irrationality here and no irrational delay; there is no failure of the agent to stop putting off decisions to preserve himself from some state he most disprefers. There is only, at worst, what we might call relative procrastination: he fails to take steps in time to prevent his arriving at a state he globally disprefers. But this is not irrational, and since procrastination is necessarily irrational, it is not genuine procrastination.

There may be a reply to this. Suppose that for each possible stopping point, the smoker with intransitive preferences weakly prefers having another cigarette to stopping but strongly prefers having smoked no cigarettes to having smoked all of them. Arguably, we can then say that there are things he prefers more strongly and things he prefers less strongly, and perhaps he should stop smoking before he gets to the state he disprefers most strongly. Perhaps then, upon surveying his intransitive preferences, he is rationally obliged to ensure that as they guide his pair-wise choices, they do not guide him to a forced stopping point that is the most strongly dispreferred of his options. He should choose as if he had transitive preferences, with that item as the least-preferred item.

But even if he has these varying strengths of preference, it is not clear that it is rational for him voluntarily to stop at any given point; for there would always be the argument that he would, however weakly, prefer not stopping until the next point, and so on. A rational person guided by his intransitive preferences would not, in fact, voluntarily stop. Indeed, this is precisely what distinguishes choosing on the basis of intransitive preferences from choosing on the basis of transitive preferences.

We have been trying to find something inherently irrational either in an agent's having intransitive preferences or in the choices that they would induce him to make. One familiar such argument is that having intransitive preferences is irrational because an agent with them can be money-pumped: regardless of which state he is currently in, he would trade some of his money to move to a different state he prefers to it, and since he always prefers another state to the one he is in, he would keep doing this until he had no more money. Surely, this makes him irrational, for there is the air of being self-defeating about him.

7. Thanks to Andreou for this objection.

OUP • UNCORRECTED PROOF

However, the money-pump argument does not by itself prove the irrationality of having or choosing by intransitive preferences. At the very least, we would have to assume that the agent has, in addition to intransitive preferences over various outcomes, a preference to keep his money. Let us assume that this is so: he will only trade his money for something he values as much or more. Suppose, further, that he prefers A > B > C > A. Suppose that he is now in state C, which we define as him having, say, a brooch and three dollars. Given a choice, he would rather spend a dollar and trade the brooch for, say, a watch; that is, he would rather be in the state of having a watch and only two dollars, which is state B. And maybe he would trade that to have, say, a pen and only one dollar, which is state A; and he would trade that to be in the state of having a brooch and three dollars, which is state C again. (Maybe someone will accept the trade, maybe not.)

Where is the irrationality? The traditional answer: in the agent's voluntarily moving to a situation that is worse (relative to his preferences) than the situation he started off in.[8] (He started with a brooch and three dollars, and now he wishes he could get back to that state.) But in fact, each position he could have been in is such that if he does not move to a different position, he is pair-wise worse off. So, he would have been irrational to stay where he was. In moving, he has not made himself any worse off than he was before. (After all, even if he made it back to the original state, he would still want to renounce it for another state, and so on.)

INTRANSITIVE PREFERENCES IN OVERVIEW AND THE IMPOSSIBILITY OF ALL-WISE CHOICES

But maybe we have not been sufficiently charitable to the idea animating Andreou's model. Perhaps the idea is this: the agent has pair-wise preferences between indulging and abstaining that always require him to indulge. But looked at from an overview, he also has global preferences according to which he would rather eschew all indulgences than engage in all indulgences. And it is this latter preference that gets frustrated by his sequence of pair-wise choices, resulting in a net irrationality. If only the agent were to take the global view, he would not engage in the indulgences but would instead act to satisfy his global preference. (One might think he would be induced to do this simply by having had occasion to take the global view, or maybe he would need to make changes to his circumstances of choice in order to prevent himself from returning to choosing by his pair-wise local preferences.

8. Thanks to Andreou for this suggestion.

OUP • UNCORRECTED PROOF

Andreou seems to think agents would have to take the latter strategy.) And if he does not take this view, he will have been irrationally myopic.

Let us see if we can represent this using a case where your preference structure incorporates global preferences, ones regarding totalities of indulgences:

$$(C) < (B \& C) < (A \& B \& C) < -(A \& B \& C) < (C)$$

Again, each letter stands for a cigarette smoked. You have the global preference for having smoked no cigarettes, –(A & B & C), rather than having smoked all of the cigarettes, (A & B & C). But now, suppose that you are asked to choose among all of the options at once, not pair-wise. Looked at from the vantage of an overview and trying to make not a pair-wise choice but an all-wise choice, you should be paralyzed, unable to make any choice, let alone one that involves satisfying your global preference. You are trying to figure out how many cigarettes to smoke, but for any number, your preferences give you reason not to pick it. You cannot pick smoking just one cigarette, because you prefer smoking two to smoking just one; you cannot pick two, because you prefer three to two; nor can you pick three, because you prefer smoking none to smoking three; nor can you pick none, because you prefer smoking one to smoking none; and so on. So, it would not be true that rationally, ideally, you should be induced by your preferences not to delay quitting smoking. For it would not be true that some number of cigarettes is the number after smoking which you should quit smoking; nor would it be true that you should stop before violating your global preference to have smoked none rather than three. And so, again, even though when offered choices only pair-wise, you will choose to violate your global preferences, there is no vantage from which this is irrational. Furthermore, since procrastination is necessarily irrational, but there is no behavior of yours that your intransitive preferences could succeed in inducing here that can count as irrational, it cannot be intransitive preferences that explain any so-called procrastination.

Note further that the paralysis you would experience in trying to make an all-wise choice from your intransitive preferences is not obviously irrational, either. For there is nothing such that, because you all-in prefer it, you are failing to advance your preferences by failing to make a choice. You are just not equipped for making choices in this situation (apart from the choice of making no choice). True, the fact that an agent's intransitive preferences allow him to make pair-wise choices but not all-wise choices has been used to argue that we cannot understand a wholly well-ordered preference ranking as just a construct of pair-wise rankings; we must add that the pair-wise rankings must be transitive. But all this means is that unless your preferences are transitive, there will be some situations in which your preferences cannot justify you in

OUP • UNCORRECTED PROOF

choosing one among several options. There is no further obligation of rationality to be such that your preferences would always enable you to make such choices.[9]

PROCRASTINATION, TRANSITIVE PREFERENCES, AND EVER-BETTER PROBLEMS

We now know that to have irrational delay from an agent in the advancing of his global preferences in the way Andreou sees it, the agent's preference structure must be transitive. There must be something he ranks highest, something he globally prefers, his ambitious goal, as well as other things he ranks beneath it, his indulgent goals, and somehow he must be led by temptation and his preference structure to put off his global goal until it is too late. In fact, we need something even more complicated: that what the agent wants most is some big, important, globally preferred thing, plus a bunch of smaller things, ones the getting of which makes him better off than just the big thing alone, but where his getting too many of the smaller things ruins the big thing. Plausibly, then, what the agent wants is as many indulgences as possible compatible with still reaching his global goal.

Let us reimagine the self-torturer puzzle to make it capture these features. If you are in the new version of the puzzle, your preferences look like this:

> maximal tolerable indulgences > maximal-1 tolerable indulgences > maximal-2 tolerable indulgences > maximal-*n* tolerable indulgences > 0 indulgences > intolerably many indulgences

But if you can tell intolerable levels—something that, as we saw, has to be true for self-torturer-like cases to be intelligible—then you can tell where to stop, namely, just before the first setting that is intolerable compared with the zero setting. But this means that there would be limits on pairwise temptations to move up; and so you would not, if guided by your preferences, fail to attain your ambition, what you globally prefer. Normally, then, there would be no procrastination, no irrational delaying of stopping before it is too late. But suppose you do not then stop. What could explain this? Not your overall preferences but only the usual explanations of weakness of will, such as distraction, confusion, and the others mentioned above.

We have been exploring problems with Andreou's view on the assumption that procrastination is modeled well by the case of the self-torturer.

9. There is, however, an obligation under certain conditions not to move from having transitive preferences to having intransitive preferences. For argument to this effect, together with a general discussion of well-orderedness in preferences, see MacIntosh, "Prudence and the Reasons of Rational Persons," especially 350.

OUP • UNCORRECTED PROOF

And it has proved problematic as a model, because its intelligibility depends on there being a threshold of disaster. It relies on subjective intolerability, and, given other stipulations about the model, without a threshold of intolerability, there can be no subjective intolerability in the model, no disaster, and so no rational objection to the agent's indulging indefinitely; therefore, his behavior could not count as procrastination. But Andreou also sees as part of the problem that when we are induced by our preferences to procrastinate, this is because there is no obviously appropriate place to stop indulging. I now turn to a family of structures with this feature. I begin with cases in which there are infinitely many prima facie acceptable stopping points, with an incentive at each point not to stop before the next, where not stopping sooner or later would result in disaster but where no given stopping point is the one proceeding beyond which is itself disastrous. I will explore whether this should present a problem for agents. I then turn to cases where, however many possible stopping points there are, they exist on a vague continuum. I shall argue that the infinite case has a solution and that the solution can be applied to the vague cases. I then explore what it would mean if I am wrong about this, suggesting that even then, such structures cannot explain procrastination.

So, suppose that Andreou had in mind "ever-better" problems, problems whose structure is nicely illustrated by the following question: On what day should you drink a bottle of wine that improves every day? Each day, there is an argument for delaying another day, for the wine will then be better; but some day must be chosen, or else one never gets the benefit of drinking. Likewise, maybe before the intolerable point in the self-torturer case, there are infinitely many acceptable stopping points, each one such that there is an incentive not to take it. Your preferences look like this:

> maximal tolerable indulgences, where the last indulgence must be chosen from pair-wise incentivized options among which there is no optimum > maximal-1 tolerable indulgences > maximal-2 tolerable indulgences > maximal-*n* tolerable indulgences > 0 indulgences > intolerably many indulgences

Let us add some structure to the self-torturer case in order to have a concrete example of this sort of ranking. Suppose that you are given the usual options of moving up a setting, and suppose also that you can determine the highest tolerable level. So, you advance to there, and now you want to stop. But you will only be allowed to stop if, in the next ten minutes, you pick a number between one and infinity. If you do not pick a number, you will be forced to the first intolerable level. Let us make the case more vivid: if you do not pick a number, you will be shot dead. You are also told, however, that whatever number you pick, you will be given its value in dollars. Can you rationally pick a number?

OUP • UNCORRECTED PROOF

For each number, there is an argument for not picking it, because it would be better to pick a higher number, since that will get you more money; and there is no highest number; so there is no uniquely rationally choice-worthy number. Thus, we can imagine an agent in this situation nominating a number, rejecting it for a higher number, and then rejecting that number for an even higher number, and so on, the net effect being the procrastination of picking a number. We would then have a case of procrastination explained much as Andreou saw it: each option available to the agent is preference-dominated by a later (in this case, higher) option, and the effect of delaying stopping the consideration of options results in a globally dispreferred outcome (in this case, death).

On the other hand, since not settling on a number is disastrous, the agent should settle on a number. But how? He could use a symmetry-breaking technique: each number is such that choosing a higher number than it is incentivized, but each number is also such that settling on it is less disastrous than not settling on any number. That is, from the point of view of averting disaster, each number is equally serviceable. And since it is more important, given the agent's values, that he pick some number than that he not pick a number (in comparison with which a higher number would get him more money), more important because he thinks it better to be alive with some money than dead with a little more, it is the latter reasoning that should control him. And this deliberative style allows him to treat all of the numbers as equally good, period, rather than just equally good in a crucial respect, for it allows him to discount the respects in which they are not equally good—if he does not discount those respects in this way, he cannot get what he most wants, namely, to keep his life.

At this point, the problem has the same logic as the one Buridan's ass faces in choosing between two equally good options. And paralysis of choice is not rational in the Buridan case. Instead, an agent in such a case should break the tie using a symmetry-breaking technique whereby a randomizing process arbitrarily nominates one of the options, thereby making it salient and therefore choice-worthy. (The arbitrariness lies in the fact that there is nothing inherent in any of the options that makes it choice-worthy; rather, the eventual choice-worthiness of one of the options derives from the fact that a process, designed arbitrarily to pick some option or other, picked it, a process whose use is justified independently of the properties of the objects among which it picks.) In a tie between two equally good options, one might flip a coin; in a tie between infinitely many equally good options, one might use a random-number generator. Perhaps one has such a generator in one's head, which one activates simply by thinking, "I shall now pick the first number that pops into my head." At any rate, we should be able rationally to do something like this if we can generally

OUP • UNCORRECTED PROOF

use symmetry-breaking techniques to choose among things we are tied about, which we can.[10]

Is this satisficing?[11] No; satisficing is forgoing an available optimum in favor of something suboptimal that is good enough. But in our case, there is no optimum to forgo. So, we have an agent who has found it maximizing to take the vantage of avoiding the worst; and looked at from that vantage, the case presents him with many equally good options, choosing among which must be done with a symmetry-breaking technique such as random choosing.[12]

Of course, it is possible that the agent's choice in this and similar cases should not be purely random. Often, there are constraints from our other goals that should affect the choice of stopping point in a given continuum.[13] For example, there are many opportunities for me to quit smoking, but they are not all equally good. Quitting just before a conference, for instance, would make it impossible for me to perform at the conference, so I should quit only after. And if I were in our new version of the self-torturer scenario, maybe I would have a rough idea of how much money would meet my foreseeable needs and would believe that getting very much more money would cause me other problems—attract criminality, ruin my personal relationships, deflate the value of the currency, and so on. So, I should randomly choose only among items in the remaining range. Of course, if there is a highest number in this range, I should pick that.

But perhaps what counts as in that range is vague. And Andreou is interested in scenarios that feature vagueness, so let us proceed to these.

CHOOSING WHERE TO STOP IN VAGUE CONTINUA

One problem agents face in Andreou's scenarios is that it is not clear what counts as too many indulgences when trying to maximize the number of

10. Things are more complicated if you and I must coordinate by agreeing on one of some many options among which each of us is indifferent (or slightly dissimilarly incentivized), each of us using our own symmetry-breaking technique in proposing to the other agent which option to settle on. I argue in MacIntosh, "Buridan and the Circumstances of Justice," that since there is no guarantee that our several symmetry-breaking techniques will nominate the same option, and since an infinite regress may begin in trying to use symmetry-breaking techniques to solve *that* problem, there is no guarantee of a rational solution to this coordination problem. I thus endorse Andreou's invoking coordination problems to explain why people procrastinate—or at least delay, for it is not clear that there is irrationality here—in dealing with issues such as pollution, where agents would have to coordinate to solve the problem; see Andreou, "Environmental Preservation." In fact, I raise her worry (in the poker sense of "raise") and claim that there may be no rational solution to these *n*-party coordination problems, barring certain lucky events that themselves cannot be rationally coordinated for by agents seeking to coordinate.

11. Thanks to Elijah Millgram for the question.

12. Sorenson, "Originless Sin," and its associated literature may be relevant here.

13. Thanks to Sue Sherwin for pointing this out.

OUP • UNCORRECTED PROOF

them compatible with avoiding disaster for one's global preferences. Having one more cigarette will not be the difference between getting cancer and not, although having many more will likely make the difference (or at least some significant difference), even though there is no definite number (or at least no known one) that will do it. This is likewise true for one more act of pollution and catastrophic global warming or one more night of watching TV rather than working on your book. If we think of the indulgences as arrayed in a sequence on a continuum, these cases have the following structure: At the start of the sequence is a vaguely bounded range of indulgences that you could engage in with no significant risk to your larger ambition. This vaguely transitions into a range where the more you indulge, the less likely it is that you will be able to attain your ambition, but it is in decreasing degree likely that were you to stop indulging, you could still attain it. And this, in turn, vaguely transitions into a range where it is very likely too late to attain your ambition, so that further indulgences do not significantly worsen your chances at it, and there is little point to stopping.

For these situations, there are several possibilities. First, perhaps you can assign utility values to attaining your ambition and to each indulgence, and for each incremental indulgence, you can determine the degree to which it reduces the probability of your being able to attain your ambition. We then have a straightforward problem in expected utility theory: you must maximize your individual expected utility, and there will be some indulgence that is the last one that increases rather than decreases it, a point at which the increasing utility from further indulgence is canceled by the decreasing likelihood of attaining your ambition, an ambition with a high and known utility.

A second possibility is that you do not know precisely the utilities of the indulgences and the ambition and/or the probabilities of attaining the ambition given each further indulgence. But you know the odds of each thing having any possible utility it could have for you and the odds of each indulgence having any possible effect on attaining your ambition that it could have. So, again, you can do an expected utility calculation by first multiplying your estimates of the utilities and probabilities by your estimates of the likelihood of the correctness of your estimates, giving you a probability-weighted value for the utilities and probabilities, which you then multiply together for each choice of indulgence, stopping indulging after the last indulgence that increases your expected utility.

A third set of possibilities is that either the probabilities or the utilities or both are such that each can be specified only within a certain limited range of accuracy. Perhaps you know that the utility of attaining your ambitious goal is somewhere between 100 and 120 happiness units or that the probability of the next cigarette's reducing by 1 percent the chance of attaining that goal is between 2 and 4 percent. Again, expected utility calculations come to the rescue: you should compute the expected utilities of each choice on each assumption of the utilities and probabilities

OUP • UNCORRECTED PROOF

in your accuracy range, average them together, and then keep indulging to the point of highest average expected utility. (Admittedly, these calculations are becoming pretty daunting in practical terms for ordinary agents, who lack actuarial acumen, but the calculations afford—in principle—a solution to the problem.)

In all three cases, procrastination would consist in indulging beyond the stopping point dictated by the varyingly complicated expected utility calculations. But to explain why you would so indulge, we would need the usual explanations from the weakness-of-will literature—it would have nothing to do with intransitivity in your preferences or even with vagueness.

A fourth possibility is that while there is a utility-maximizing choice, you do not know it, and you are aware of this; you know only that the longer you indulge, the more money you get, but the less likely you are to attain your larger goal, and that this goal is more important to you than any amount of money. Here, you should take no chances and should stop indulging immediately; you should take the step certain to avoid disaster. And were you to fail to do this, the explanation could have nothing to do with intransitivity in your preferences or with vagueness.

A fifth possibility is that you face inherent vagueness, itself at best only crudely specifiable. Imagine a variant of the self-torturer case where you must satisfy an administrator before you are allowed to stop raising the shock levels. For example, suppose that in order to exit the self-torturer scenario, you must watch the administrator dropping grains of sand onto a table, and you must tell him to stop before he has made a heap, but you will also be given a dollar for each grain you let him drop.

Here, the difference between something's not being a heap and being a heap will not be one grain of sand. Nevertheless, there will be some clear cases of a nonheap and some clear cases of a heap. Your job is to allow grains of sand to be dropped but only to the point beyond which it would be strongly arguable that there was a heap. Unfortunately, there is no precise such point for heaps, but there is a range of points within which all of the points are such that there is no argument whatsoever for thinking that there is yet a heap—call this the safe range. Of course, the safe range itself has vague boundaries, and likewise for the continuum of any vague concept.

Must an agent having to choose here be pulled into procrastination? No, because he should be able to think this way: no criteria internal to the concept of the vague thing in question can guide the fine-tuning of where to draw the line. Suppose that you are our self-torturer and that the administrator is playing fair; he will be as reasonable as you in drawing lines, and he must accept that, provided you do not misdraw a line by labeling a case that is clearly one thing as clearly another, you have chosen permissibly, and you will be allowed to stop raising the shock levels. It is rational for you randomly to choose a point within the vaguely bounded safe region, a point toward its later extreme in order to get more money. The same applies to deciding how many more cigarettes to smoke. Doctors think that smoking for 20 years strongly increases cancer risks, smoking

OUP • UNCORRECTED PROOF

between one day and five years not so much. So, it is within that range, roughly, that one should make a random choice of when to quit (see the concerns I discussed earlier about when exactly to quit). The same goes for how many more days you can delay enacting pollution legislation or how many more evenings you can laze around failing to work on your book.

An agent rationally should not be pulled into procrastination by vagueness. Of course, I concede to Andreou that if the agent deals with vagueness irrationally, and so procrastinates, her procrastination will have been caused in part by vagueness. But since both the rational and the irrational person are contending with vagueness, it must be something other than vagueness that explains the difference between them—that explains the irrationality of the irrational person.

IMPLICATIONS FOR THE POSSIBILITY OF PROCRASTINATION AND ITS EXPLANATION

We have considered many forms of choice problems, posed to agents with either transitive or intransitive preferences. But for each problem, there are only (coincidentally) five possible implications for the possibility of procrastination and its explanation.

First, there may be a rational solution to the problem of where to stop indulging, a solution dictated by the agent's preferences and beliefs about the circumstances (including his mastery of the concepts in play in the event of vagueness), and the agent knows it and would therefore take it, in which case we will not have procrastination. If he does not take it, then his behavior is not properly under the control of his preferences, in which case, in opposition to Andreou, his preferences are not the problem.

Second, perhaps there is a rational solution, but the agent does not know it and therefore will not take it. But then there is still no procrastination, which is defined as irrational behavior; as long as the agent has the excuse of ignorance about which solution is rational, he cannot be called irrational for failing to take it, at least not unless he is culpable for not knowing. But if he is so culpable, as a result of wishful thinking or whatever, then the procrastinating is being caused by one of the usual suspects in the explanation of weakness of will, and, in opposition to Andreou, we have no special explanation for procrastination in preference intransitivity or vagueness.

Third, perhaps there is no rational solution. (For example, maybe some of the proposals I made above about what would be rational solutions to some of these problems fail, and maybe no proposals could succeed.) But if there is no rational solution, then neither is there such a thing as culpably failing to take it. So, again, the agent is guilty of no irrationality and is not guilty of the form of it which is procrastination.

Fourth, it could be that what counts as a rational solution is inherently vague. But then it is, at worst, vague whether the agent is irrational, at

OUP • UNCORRECTED PROOF

least if his choice fell in the vague area of arguably rational but also arguably irrational solutions. If it fell in the clearly irrational range, again, we would have to invoke one of the usual suspects cited in the explanation of weakness of will—nothing special about procrastination.

Fifth, maybe some of these problems, such as those we considered earlier for agents with intransitive preferences, have prudent delay as their solution. (For example, if one has intransitive preferences over the number of cigarettes to smoke, one should keep smoking more cigarettes, which is to say that one should delay quitting smoking.) But then, again, an agent taking this solution will not be guilty of procrastination, for she will not be guilty of any form of irrationality.

It appears, then, that, in spite of the plausibility and elegance of Andreou's conjectures, neither intransitivity in an agent's preferences nor vagueness in what counts as an appropriate point to stop indulging in favor of attaining larger goals in the course of advancing transitive preferences can explain procrastination conceived as irrational delay. Andreou correctly sees that an agent with intransitive preferences, for example, concerning how many cigarettes to smoke, will delay quitting smoking, resulting in his not satisfying his global preference to have smoked no cigarettes rather than many.

But since his preferences are intransitive and since there appears to be nothing irrational about having such preferences or about making local-preference-advancing pair-wise choices from these preferences—choices to keep smoking one more cigarette—Andreou cannot represent the agent's failing to satisfy his global preference as irrational. She can explain his delay in failing to stop smoking in time to satisfy his global preference but not the supposed irrationality of the delay. Andreou is also right to see that vagueness, namely, in what counts as an appropriate place to stop indulging in pleasant activities that delay the attainment of more important goals, can pose difficult choice problems for agents with transitive preferences. But if, as I have suggested, these problems can be rationally overcome with symmetry-breaking techniques, then, if an agent fails to overcome the problems in this way, the explanation must lie elsewhere than in the vagueness regarding when it is best to stop indulging. We must appeal to things such as weakness of will to explain the agent's failing rationally to solve his problem. Meanwhile, if I am wrong about there being a rational solution to the vagueness problem, then an agent who fails to solve it cannot be accused of any irrationality, for there would be no rational course that he has failed to take. And so his failure could not amount to procrastinating, for it could not count as irrational delay.

ACKNOWLEDGMENTS

For useful conversation, my thanks to Steven Burns, Sue Campbell, Richmond Campbell, Carl Matheson, Chris Olsen, Susan Sherwin, Heidi

OUP • UNCORRECTED PROOF

Tiedke, Michael Watkins, Sheldon Wein, and a colloquium audience at Dalhousie University. For written comments, my thanks to Darren Abramson, Chrisoula Andreou, Bob Martin, Mark White, and Greg Scherkoske. Thanks also to the participants in the workshop held for this volume, particularly George Ainslie, Olav Gjelsvik, Elijah Millgram, Sergio Tenenbaum, and Frank Wieber. Finally, thanks to Gjelsvik and to Jennifer Hornsby for their sponsorship of and their work in putting together the workshop and to Andreou and White for their organizational and editorial efforts.