

## Shifting the Problem of Akrasia

**Nathaniel Kastl**

Texas A&M University-Corpus Christi

### **ABSTRACT**

The Problem of Akrasia is a longstanding paradox that arises when decision making is considered to be a chiefly rational process and a person is expected to make the best possible choice for themselves (as rational people can sometimes be observed to knowingly choose the worse of two actions). This paradox presents a serious problem for most models of decision making. Recently, Donald Davidson has offered several promising solutions. In this paper, Davidsonian thought on the issue will be thoroughly investigated and ultimately found unsatisfactory. Instead, a new 'hedonistic' model of decision making will be proposed based on conclusions drawn from Davidson and Aristotle's work on the issue. In this new model, the Problem of Akrasia does not emerge.

### **KEYWORDS**

Akrasia, Donald Davidson, Aristotle, Decision Making

## compos mentis

A particularly handsome friend of mine had been having a eudaimonius spring semester when one of his exes reached out to see if he would come over. Now, this friend wasn't really one to be indecisive or easily confused, so it wasn't long before he came to a simple conclusion: She was bad news. After calmly thinking it over, he decided that he should text her back and tell her that he was not interested. A few minutes later, he found himself on the way to her house to do the exact opposite. Naturally, the next few days were a whirlwind that left him exactly as humiliated, emotionally battered, and angst-ful as he had predicted.

So why would this very-distinct-from-me person have made such a terrible decision for himself when he had already correctly predicted the outcome and determined that it was undesirable? This question already assumes much, but it can safely be said that my handsome friend was experiencing firsthand what philosophers have referred to as the Problem of Akrasia.

Though not the coinage of the term or the earliest commentary on the issue, a sort of de-facto early codification of the Problem of Akrasia can be found in Aristotle's *Nicomachean Ethics* (Book 7). Within, Aristotle describes and attempts to solve the observed paradox present in human behavior in which a person appears to rationally select the best course of action for themselves, then fails to act on it. Importantly, Aristotle goes on to make the following exceptions: A person cannot be said to be exhibiting akrasia if their consideration failed to include all relevant information, or if their consideration was logically flawed, or if their considerations were overwhelmed by their passions, or if the person was in any other way indisposed to proper rationality. In these cases, Aristotle is quite comfortable with the a priori explanations for their behavior. For the cases that remain, however, he offers two striking propositions:

1. Decision making is chiefly rational (outside of the aforementioned exceptions)
2. The Socratic Assumption: or No person knowingly and willingly makes the worse of two choices for themselves

Both of these propositions seem quite compelling, but observational (and anecdotal) evidence would seem to suggest that they contradict each other as together they paint a picture of reality that is simply not the case.

To solve the Problem of Akrasia, Aristotle suggests that the first premise is flawed. Instead of being the center of decision making, he suggests that rationality might be but one of several competing factors, and that this puzzle in particular might be accounted for via another factor for decision making: a desire for pleasure. This conclusion is somewhat distasteful to Aristotle as in the wider context of *Nicomachean Ethics*, it indicates that having great faculties of rationality does not actually guarantee that a person will not commit wrongdoing. To Aristotle, the only real defense that remains is to attempt to quell one's own desires through habit. If I had asked him to explain the actions of my friend, Aristotle probably would have told me that his rationality had been beaten that night by his lazily disciplined set of habits- and then changed the subject.

### **A MORE CONTEMPORARY APPROACH**

Of course, the work of Aristotle is beyond steeped in what could generously be called the eccentricities of his time and lacks the more thorough structure and extensive collaborative processing of our time. In the modern era, one name is seminal with work on the Problem of Akrasia: Donald Davidson. In 1969, Davidson fundamentally altered and modernized thought about the Problem of Akrasia. Not only is his work a practical well of information and argumentation that evolved dynamically across his career, but it also serves as a hub of discussion for the philosophical community at large.

Davidson's account of the Problem of Akrasia is refreshingly distinct from the Aristotelean thought that came before it. In particular, he insists that the Problem of Akrasia is poorly served by being viewed through an ethical lens, as supposedly akratic situations can be amoral just as often as they can be within the scope of ethical consideration, like in the case of a person who acts inconsequently against their own intention, or even have morally superior outcomes, like in the case of a person who believes an evil act would be the best course of action for themselves but refrains akratically. Additionally, Davidson resists the Aristotelean position that the problem might be solved by attempting to manipulate the premises that created it.

Davidson lays out his account of the problem most clearly in his 1969 paper *How Is Weakness of the Will Possible?*. Davidson presents three premises:

## compos mentis

1. If a person believes one course of action to be better than a second course of action, they *want* to take the first
2. If a person wants to take one course of action more than another, then they will take that course of action if they can only take one
3. Sometimes, a person will not take the course of action that they believe to be better.

Like Aristotle's, these premises are all compelling but seem to be contradictory. Unlike in the case of Aristotle, however, Davidson does not attempt to solve the problem at this level. Instead, he re-imagines what (even in his argument) is considered to be an akratic action. In this new understanding, the person acting contrary to their judgement is in fact not ignoring the parts of consideration involved in the first two premises (the selection of the best action and the transition from an opinion of a best action to a desire to take that action), but are rather performing those considerations hastily. In other words, a person who acts akratically has indeed selected a course of action that they believe to be better, has wanted to perform it, and then has performed it- only they did so without fully considering the wider context relevant. This phenomenon only seems to be a paradox because that same person can also perform the same consideration to a greater, more complete degree at a different time (before or after the fact) and as a result deem the action taken to have been against their better judgement.

Notably, this solution bears a clear resemblance to one of the Aristotelean exceptions to true instances of akrasia (the event of a consideration that lacks all relevant information). As a result, it can probably be said that Aristotle would not consider this to be a solution to the Problem of Akrasia at all, but rather a claim that what appears to be akratic actions are in fact just a type of action that he has already dismissed as unremarkable. That's not to say, however, that Davidson is incorrect or even that this observation is an objection at all. In fact, the explanation of akratic actions as simply being actions of another less interesting type is a very common theme in contemporary thought on the issue, both for Davidson in this as well as other explanations and for his peers.

Moreover, there is actually an important distinction between Davidson's account and Aristotle's exception: Davidson's account includes the secondary 'complete' consideration that causes the person taking the action to notice the issue in the first place. If this consideration was not involved in the scenario, it would indeed be the case that Davidson simply describes a person who fails to make a decision well given their capacities as accounted for by the exception. While it could certainly be argued that this, in reality, is what akratic actions are, it is not the case that this is what we *mean* when we say that we are experiencing akrasia because in this case the person performing the action, limited by the lesser amount of knowledge considered, doesn't actually know that they have done something seemingly akratic (which ostensibly is very much the point). Instead, when the wider secondary consideration is performed that does actually contain all relevant information, the person becomes aware of what might be called their better judgement and as such is then able to become aware of the fact that they have acted against it. Thus, Davidson's account is distinct in that it is representative of what is meant by the term akrasia where Aristotle's exception is not.

For a better objection to Davidson's position on the Problem of Akrasia, then, it might be more productive to refer to those he makes of himself. In his later paper, *Intending*, Davidson points out a flaw in the more granular mechanics of his explanation. Using the example of a person who 'intends' to build a squirrel house for no particular reason and never does, he introduces the concept of a 'pure intention', or an intention that cannot be connected to a consideration based on desires held at the time or a real action. This metaphysical gap puts Davidson's model in a precarious position as an unclassified sort of intention poses exactly the same problem that akrasia did in the first place: namely, that it is inconsistent with the model of decision-making being presented. In attempted solution, Davidson argues that 'pure intention' may be a sort of sibling to 'ordinary intention', which would actually be able to be identified with the consideration present in his first premise. In this way, Davidson pins 'pure intention' in place in relation to his model. Regardless, 'pure intention' lies outside of Davidson's model, connected as it may be, and this difficultly underscores a clear lack of completeness.

## DAVIDSON'S REFORMULATION

A later explanation of akrasia by Davidson may be more satisfactory. In *Paradoxes of Irrationality*, Davidson makes a much more radical attempt to explain the Problem of Akrasia that even arguably violates his rule about manipulating the original premises. In this reformulation, Davidson argues that what is often called irrationality is, in all actuality, simply normal rationality that is not completely understood. He comes to this point by first identifying the concept of a 'rational core' of an action as being the reason that is given when that action is rationalized. Because Davidson believes that all actions taken by a person can be rationalized in some way, he further argues that, though some actions may be irrational in certain other senses of the word, no actions are irrational in the sense that they lack a 'rational core'. In a sense, he argues that even if there is not a good reason, a person taking an action always has at least some reason for having done so. Thus, akratic actions based on irrationality in this sense cannot be said to exist if this model is accepted. Davidson doesn't stop here, of course.

If this is the case, he argues, then the real paradox lies within the case of a person who fairly weighs two separate actions, finds one to be superior to the other, then takes the other- and moreover, the real inconsistency is introduced due to the fact that he still maintains that a person ought to act on their own best judgement (the Socratic Assumption).

Naturally, Davidson only introduces a more nefarious problem in order to later introduce a more nefarious solution- or two, rather. First, Davidson points to a surprisingly original idea: the difficulty verging on impossibility of properly communicating one's drives and the impact on how this might affect our perceptions of others acting akratically. To home in on this point, Davidson introduces two new principles- the assumption of intersubjectivity and the assumption of interpersonal interpretation. The former, intersubjectivity, posits that a person cannot truly understand another person's beliefs, desires, or intentions unless the two individuals share (or perceive themselves as sharing) a vast amount of positions that they would categorize as common knowledge. The latter is very similar. With the assumption of interpersonal interpretation, Davidson presses further into the issue and argues that one must actively believe that the person they consider is wise and that what they hold to be true actually is true, else they will often tend to consider them to be irrational. In fact, Davidson argues, as the

magnitude of the deviation from our beliefs to another person's increases, so too does the difficulty of understanding them in rational terms, rational though they may be. As a result, much of what is considered to be akratic in the first place might simply be explained by a failure in what Davidson speculates to be a profound reliance on empathy and sympathy for others different than ourselves. In other words, since the act of rational consideration is so sensitive to its starting conditions (i.e., a person's values, beliefs, etc.), a failure to properly recognize these starting conditions as valid in others can cause a person to fail to properly recognize a fully rational consideration, ultimately manifesting in the phenomena of perceiving others as acting akratically.

Davidson's second solution to his more complex formulation of the problem (and his main thesis for *Paradoxes of Irrationality*) is that many actions judged to be irrational may have been the result of a cause that was mental in nature but was not a reason. From here, Davidson expands to a partitioned theory of the mind. The main suggestion of this theory is that the mind consists of multiple parts which can sometimes act against each other. While one of these departments consists of the reason and consideration described in his previous formulations, the other designates the opposing course of action. In this theory, though only the first part matches what is meant by the more typical verbiage of consideration, both parts exhibit their own versions of reasoning and justification based on available resources, so nothing truly akratic or irrational is actually occurring. Instead, the course of action is made based on the seizure of control of one of the two competing parts after full and fair considerations have been made with differences in the action selected only resulting from differences in the source of the consideration, not the quality of the consideration itself.

In a sense, this partitioned theory of the mind is only meaningfully different from intersubjectivity and interpersonal interpretation in that the two entities are contained by a single 'person' (though this terminology gets more tenuous here). Mechanically, it works the same: The Problem of Akrasia is sourced to the dissonance caused when a controlling rational consideration is ignored due to a failure to recognize the desires and drives behind it, whether that key consideration be performed by another person or by a portioned section of the same mind. Together, the two explanations approach complete coverage of the issue by explaining issues of akrasia both within the self and in others.

## **PROBLEMS FOR DAVIDSON'S LATER SOLUTIONS**

Of course, these explanations come with their own sets of issues. Intersubjectivity and interpersonal interpretation share the same problem as Davidson's first solution to the Problem of Akrasia- that they simply identify perceived akratic actions as a more mundane, known sort- but to a much more critical degree. Even if the scope of explanation for this solution is limited to akratic action in others, it cannot be said to fully achieve that goal. By saying akratic actions in others are merely our misunderstanding, one neglects the fact that the very interest in the Problem of Akrasia in the first place is in the confounding personal experience of it. A person who perceives themselves as having behaved akratically surely cannot be said to be doing so as a result of not being similar or sympathetic enough to themselves, and shelving momentarily the Problem of Other Minds, it would stand to reason that since the Problem of Akrasia is a widely reported phenomenon that seems to just be a part of being a person, this intersubjectivity and interpersonal interpretation resistant explanation cannot apply fully to what surely is the very same experience of akratic action in others from their own perspectives. Though some actions perceived as akratic in others may very well be simply a result of failure to understand their motives, this explanation rhetorically fails to demonstrate that all actions perceived as akratic in others are as such.

The partitioned theory of the mind does fill in this blind spot, but it has its own difficulties- mainly that to suggest such a thing would carry enormous ontological weight. Really, other than the fact that it seems to work as a solution to the Problem of Akrasia, there is little reason to think that the mind works in this way at all. It could just as easily be the case that Davidson's sealed-away consideration is instead just forgotten, or that an unknown part of the mind vies instead to select the lesser of two choices; explanatory power is an insufficient justification that becomes more egregious the more complex and counterintuitive a theory becomes.

So why, then, did my friend make such an awkward mistake in his love life? Were his faculties of reason overcome by a non-fit-of-passion desire? Was he just hasty and limited in his consideration? Did a secret part of his mind make the decision for him? Ultimately, each of these explanations are convincing only to a limited extent. Wherever Aristotle left his model of decision making too vague,



Davidson seems to have made the opposite mistake, constantly bloating into over specificity and leaving behind gaps due to a too-limited scope. Perhaps even more damning is the fact that Davidson's insistence that the solution to the Problem of Akrasia lies outside of the premises that devise it limits him entirely to argumentation that tries to recontextualize the phenomena after the fact.

Frankly, his reason for doing this is mystifying. The classic approach to addressing a paradox, which the Problem of Akrasia certainly is, is by carefully examining the structure of the argument and rejecting at least one premise or part of the reasoning. For this particular paradox, pointing to the reasoning as the issue can be more or less eliminated from consideration since the reasoning is so brief and simple, so the only clear approach that remains is to reject a premise. To his credit, Aristotle does address the Problem of Akrasia in this way when he concedes that decision making might not be a purely rational process. And, to his credit, the formulation of the Problem of Akrasia that appears in Davidson's *How Is Weakness of the Will Possible?* attempts to account for this adjustment with the addition of a second premise that converts a selection by reasoning to desire.

But why stop there? Surely Aristotle and Davidson's concessions both reveal that desire is a necessary component for modeling decision making. Conversely, reason does not seem to have earned its place in the model yet. In fact, as Haidt argues in *The Emotional Dog and Its Rational Tail* (2001), psychological evidence would seem to suggest that reasoning in the sober sort of way considered by Aristotle and Davidson can actually *follow* the decision-making process rather than preceding it. Haidt's evidence further indicates that what one might take as being sober consideration is actually justification and is in no way involved with real decision making but is believed to be to reinforce the illusion of self-rationality, similarly to how optical illusions and fabricated memories work by preying on the parts of our brain that already lie to us and justify our behavior. In other words, the insistence that decision making include rational consideration could very well be a mistake of correlation for causation with only the Problem of Akrasia to serve as a counterexample that occupies those cases where the correlation fails in a way that causation never would be able to. The fact alone that the Problem of Akrasia even persists indicates that a rational formulation of decision making is a poor fit and leaves desire alone in being a known component to the process.

## A HEDONISTIC MODEL

So, lacking a rational consideration-based model, what would a desire based or 'hedonistic' model of decision-making look like? Simply put, this model would posit that in any given moment, a person acts on their greatest desire. This model is not a new one, nor are the most basic objections against it new. One may argue, for example, that under this model a person would never work out, get vaccinated, or otherwise subject themselves for discomfort for a future pay off since only the greatest desire in the present is considered. The response to this objection, of course, is that a desire held at a particular time can still apply to the future or involve planning. A person being vaccinated does certainly have a desire not to receive an arm pain, but this desire is outweighed by a greater desire (still in that moment) to avoid the greater discomfort of becoming ill. Similarly, the objection that people would act immorally (stealing or resorting to violence) if they only acted on their immediate desires is addressed by pointing out that there are certainly desires that could outweigh any desire to do wrong, like a desire to be moral or a desire to avoid the consequences of violating norms. Only when these desires are absent or outweighed by other desires are immoral actions possible, as reflected in real life.

When the Problem of Akrasia is approached with this hedonistic model, it simply ceases to be a paradox. In familiar Davidsonian form, akratic action would be modeled thusly:

1. If a person desires to take one action more than another action, they will
2. Sometimes, a person takes an action that is different than the action that they have formulated a rational argument in favor of

Both premises are compelling, but this time, they do not seem to be in conflict. The solvency of this model goes deeper, though. Unlike Davidson's *How Is Weakness of the Will Possible?* solution, this solution in no way attempts to argue against the existence of akratic actions. While Davidson argues that akratic actions are actually a sort of failure of rationality (akratic actions are the result of hasty consideration, so not fully rational), the hedonistic model still allows for

akratic actions (in that a fully rational consideration is made and disregarded) while still being internally consistent. Even Davidson's issue with the placement of 'pure intention' is easily resolved via the consideration that a desire (say, to build a squirrel house) could just be relatively weak and often outweighed. A person could have a pure intention in that they desire to one day build a squirrel house, but on any given day they desire some other activity more. This seemingly unimportant detail is actually key evidence that the hedonistic model succeeds where Davidson fails as it doesn't suffer from the hyper-specificity that excludes certain components of decision making.

On the other hand, perhaps the hedonistic model is so intuitive that it wraps around and collides with the Problem of Akrasia in a different way. I'm sure my handsome friend would certainly not be convinced that he did not have a very strong desire to continue having a focused and flourishing semester and another desire to avoid heartbreak- surely these weren't overpowered by a desire for a relationship he was consciously suspicious of. It would seem to be the case that he believes (based on his desires, of course) that he ought to have acted in a certain way but then acted in a different way. There are two important ways to respond to this objection. First, and most uncomfortably, is the implicit suggestion contained in the hedonistic model of decision making that pushes so many away: that we might tend to delude ourselves or be ignorant about what our real (and acting) desires actually are. The meaning of this suggestion ranges all the way from the concept of purely subconscious desires that manifest in confusing ways to the idea that we might simply be less than honest with ourselves when we desire something that we wish we had not. In the case of the dieting man who is shocked to find that he has eaten a sundae, there is a simple and tactless response: he probably just wanted to eat that ice cream more than he wanted to lose weight. Though this may seem profoundly different from the idea of purely subconscious desires, they really amount to the same thing, which is to say they are both reasonable enough explanations that cannot really be engaged with or debated because they are entirely reliant on discussion about subjective experience.

Relying on this sort of argument is unsportsmanlike in conversation and bad form in philosophy, but that's not to say that it's incorrect. The limitations of subjectivity do stop (productive) conversations in the case of a disagreement, but they don't define the truth quality of statements made in the technical unknown.

Luckily, a much more functional argument can be made to explain the disconnect between how a person believes their desires are arranged and the desires that they seem to act on. In the case of a desire that is clear and present, like a sundae or a short drive to meet an old flame, there is little doubt that acting accordingly to the desire will result in said desire being realized. However, in cases of delayed gratification, like good nutrition or one's long-term mental health, the same cannot be said. As a result, it may well be the case that these two different sorts of desires are treated differently within the hedonistic model- either by actually being of different qualitative sorts, or simply by being of different relative weights according to their immediacy. With this in mind, akratic actions are even easier to explain under the hedonistic model. To return to the case of the man with the sundae, it might be more tactful and more accurately explanatory to suggest that his known desire to eat the sundae won out because it had a more immediate outcome than his desire to be healthy. Notably, though, less immediate desires can still outweigh immediate desires given appropriate magnitude. If the sundae offered to the man came at the cost of a prison sentence starting a year after he ate the sundae, he would clearly leave it be.

Another major objection to the hedonistic model solution lies within the potential hypocrisy of the proposal. Surely it cannot be the case that Davidson's partitioned mind theory fails because it is only supported by its solvency of the Problem of Akrasia while the hedonistic model does not suffer the same critique. In all actuality, this is a real problem for the hedonistic model. In defense, one might argue at the very least that the hedonistic model has the benefit of being vastly more intuitive than the partitioned mind, which sounds frivolous but is poignant when considering the comparative burdens of proof. Additionally, the evidence offered by Haidt would seem to suggest that the rational basis on which the partitioned mind theory relies is flawed in such a way that the hedonistic model is not, nearly to the point of implication. The only real defense to this criticism, though, is Occam's razor. As problematic as it is to posit theories that exist only as a response to problems and not as an extension of observation of reality, it is an essential part of how philosophy and science both progress. Occam's Razor is an important tool for controlling this sort of theory that states that when possible, the simpler of two equally effective explanations is more likely to be reliable. Under Occam's Razor, the preference from theories like the partitioned mind to the hedonistic model is undeniable.

What about our intuition that we do have rationality, though? Do we really never act according to our rational considerations and just justify backwards? Fortunately, this question does actually have a reasonably satisfying answer baked into the hedonistic model itself: We desire rationality. If a person desires to be rational, as many do, the weight of a desire in favor of choices supported by one's reason increases. This caveat is hardly enough to brighten the outlook of such a grim model, but it does offer at least a ray of hope for consideration. In a way, just opposite to how the Aristotelean model characterized decision making as being mainly rational with components of desire factoring in, the hedonistic model is mainly based on desire, but still contains a component of rational consideration.

Is this model enough to explain what happened to my friend all those weeks ago? It's bizarrely satisfying to give up a bit of control and allow oneself to believe that actions might not be the result of thoughtful consideration, most of the time or ever. I'm sure my friend still isn't completely comfortable with the idea that what he thought he wanted wasn't really what he wanted, or that just that the timing of events was enough for his desires to manifest in such a way that brought him to harm, but a solution to the Problem of Akrasia was never going to bring comfort. For all my friend can know, that very same evening may well play out again, even despite his best intentions and loftiest thoughts about himself. One almost wonders if the reason that the Problem of Akrasia has remained unsolved for so many millennia is because the real comfort is in putting a name to the unknowability and lack of control that we find even within ourselves.

## REFERENCES

- Aristotle. c.335. *Nicomachean Ethics*. Translated by Christopher Rowe. Oxford: Oxford University Press.
- Davidson, Donald. 1969. "How Is Weakness of the Will Possible?" *Moral Concepts*.
- Davidson, Donald. 1978. "Intending." *Philosophy of History and Action* 11: 41-60.
- Davidson, Donald. 1982. "Paradoxes of Irrationality." Oxford: Oxford University Press.

compos mentis

- Gosling, Justin. 2002. *The Weakness of the Will*. Abingdon: Routledge.
- Haidt, Jonathan. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (4): 814.
- Hurley, Paul. 1992. "How Weakness of the Will Is Possible." *Mind* 101 (401): 85-88.
- Mele, Alfred R. 1983. "Akrasia, Reasons, and Causes." *Philosophical Studies* 44 (3): 345-368.
- Saarinen, Risto. 1993. "John Buridan and Donald Davidson on "Akrasia"." *Synthese* 96: 133-153.
- Tremblay, Michael. 2020. "Akrasia in Epictetus: A Comparison with Aristotle." *Apeiron* 53 (4): 397-417.