

Explaining the Virtue of Consistency

John Symons

Forthcoming in *Filosofia Maravilhosa: Journal of the Brazilian Academy of Philosophy*

1. Introduction

Finite agents whose minds have the complexity and cognitive resources sufficient to engage in a normal adult human life seem doomed to live with inconsistency. We are inconsistent in our beliefs, in our pronouncements, and in our preferences. Given, for example, that one is not logically omniscient, and given a large enough set of beliefs, the goal of knowing that one has achieved consistency will generally be unattainable in practice.¹ Nevertheless, we still care about *trying* to be consistent. We value the virtue of consistency and see it as somehow closely connected to rationality, agency, and personal identity. The moral and practical role of consistency elicits a number of related conceptual questions: How and why should consistency motivate us? In what ways, if any, is it constitutive of our agency? And are there distinctive reasons to be consistent that are not reducible to other goods?

Kant put consistency at the core of his account of agency and morality. His categorical imperative (in at least two of its formulations) makes consistency a central criterion for evaluating the moral status of an action.² However, even Kant's most dedicated supporters recognize the difficulties associated with grounding claims about agency and action on a purely formal notion like consistency. Kant's view on this matter is widely regarded as presenting what Onora O'Neill calls "a dismal choice between triviality and implausible rigorism" (1985, 183). But is there a way to understand the role of consistency that does justice to its normative role while situating it properly within a realistic picture of fallible and limited human agency? I will argue that there is and in this paper I will offer an emergentist explanation of the place of the epistemic virtue of consistency in the context of social institutions and practices. Ultimately, the definition I will defend here involves an explanation of the relationship between the formal character of consistency and the virtue of

¹ The problem of logical omniscience is most frequently discussed in the context of criticisms of formal approaches to belief and decision making like epistemic logic or decision theory. Putting the worry formally, critics pointed out that the relational semantics of epistemic logic commits one to a closure property for an agent's knowledge that is implausibly strong given actual human reasoning abilities. The closure property gives rise to an objectionably powerful kind of logical omniscience:

Whenever an agent c knows all of the formulas in a set Γ and A follows logically from Γ , then c also knows A .

What this means is that c knows all the logical consequences of any formula (or set of formulas) that c knows. Logical omniscience is clearly too much. It would be a feature of unlimited rationality and it would be equivalent to superhuman abilities like the power to know all the tautologies. Early critics of epistemic logicians argued that because it resulted in bizarre consequences of this kind, the mathematical formalism of epistemic logic is simply unsuited to capturing actual knowledge and belief as these notions figure in ordinary human life. For more on this see Rendsvig et. al (2019).

² For a defense of Kant's universality tests that emphasizes the centrality of consistency, see O'Neil (1985).

attempting to act in accordance with formal consistency. To exhibit the virtue of consistency is to exhibit a disposition to shape our actions in a way that we recognize as bearing some resemblance to the formal principle. Specifically, as I will argue, it means to be disposed to correct oneself in such a way that one tends to act in accordance with the principle that we ought to be consistent in our beliefs, utterances, preferences, and actions.³

Consistency, rationality, and other epistemic virtues are closely connected and it is easy to confuse them in ways that make these notions harder to explain. In recent decades, philosophers have begun to untangle some of the related strands in the explanation of the epistemic virtues. We have recognized various distinctions that were often blurred in traditional discussions; we recognize, for example that being rational is not the same as being consistent, that open-mindedness is not the same as curiosity or conscientiousness, etc. Philosophers are developing a more complicated and accurate picture of the epistemic virtues and are steadily filling in the details (Turri et. al 2017, Horgan et. al 2018). This project is not just about achieving descriptive accuracy and completeness, it also helps answer traditional philosophical questions and to provide better explanations. As we shall see, one can read Niko Kolodny's well-known arguments against the project of giving reasons for being rational as closing off one relatively sterile explanatory project while allowing us to refocus on distinguishable aspects of human rationality rather than on the less tractable notion of rationality *per se*. The epistemic virtue of consistency is one of these distinguishable aspects of rationality that is amenable to philosophical explanation. We can argue over what explanation of the virtue of consistency is correct, but there is no persuasive argument to the effect that such explanations are *impossible*. Thus, a divide and conquer strategy for explaining the epistemic virtues looks promising to me.

The account I provide here assumes that the epistemic virtue of consistency should be understood as playing a corrective role (Alsharif and Symons 2021). However, the corrective role of the virtue of consistency only makes sense within emergent social institutions. In characterizing consistency as a corrective virtue, I am following Philippa Foot's approach to the moral virtues but unlike Foot, the present discussion is oriented less to the inherent weaknesses or vices of what she considers our natural human condition and more to the nature of our social institutions. Foot defends the view that the moral virtues are correctives that respond to deficiencies in "ordinary human nature" (Foot, 2002).⁴ On my view, the virtue of consistency serves as a corrective in ways that make sense only in very specific social contexts.

³ Thanks to Erin Frykholm for emphasizing the importance of the distinction between *in accordance with a rule* and the determination of an action by a rule. Assuming that our actions are motivated by contradictory sets of beliefs and preferences, then the best that the corrective virtue can do is to bring our unruly nature into some degree of accordance or conformity with a formal rule. In this sense, the consistent person is not following or determined by the rule.

⁴ For instance, courage is a virtue only insofar as people are ordinarily inclined to flee or submit in the face of dangers when they ought to stand firm. If human beings were not subject to such inclinations, courage would not be a virtue for us. See Alsharif and Symons 2021 for a discussion of open-mindedness as a corrective epistemic virtue in this sense. The argument of that paper is that open-mindedness is a corrective to the negative effects of confirmation bias.

The virtue of consistency not only plays a corrective role, it is also partly constitutive of certain kinds of social agency in ways described here. As explained in detail below, a condition for the possibility of continued participation in social institutions that provide access to distinctive and important goods is that an agent aim for consistency. In other words, one must aim to be consistent in order to have enduring access to the kinds of agency and their related goods that are associated with participation in these institutions. What I mean by social institutions here is very inclusive. For example, one's ongoing participation in conversations, in loving relationships with other persons, in education, and in markets, will all depend on a commitment to consistency.

The easiest case to explain is market agency and behavior. As we shall see, one's continued existence as a market agent depends on one's being guided by the virtue of consistency. In this way, the virtue of consistency is partly constitutive of what it means *to be* a market agent. If one consistently fails to be consistent, one puts one's continued existence as a market agent at risk. While I concentrate on market behavior here, other kinds of institutions and agency have similar features and similar stories can be told for them.

What does it mean to provide an *emergentist* account of an epistemic virtue?⁵ Quite simply, we begin with the observation that there is no virtue of consistency in a world without the relevant kinds of demands or conditions that require agents to be consistent. These demands or conditions emerge along with their associated virtues at some period over the course of natural or social history. On this view, at least some of the virtues are sustained by, and make possible the existence of co-emergent social orders. This is not to say that consistency is *a human invention* or *a social convention* in the sense that communities agree on or invent the logic underlying consistency. That clearly cannot be the case. Indeed, as a matter of logic, and in a sense much more basic than convention or agreement, all behavior requires that an organism behave in ways that do not frustrate its own behavior. From this perspective, once minimal kinds of agency appear in nature, far earlier than the emergence of societies or human communities, we can detect a place for something we would call consistency. However, notice that prior to the appearance of communities and institutions of the right sort, consistency should not be regarded as a *virtue*. Simple biological systems simply die or die out if they are unable to behave in ways that facilitate their metabolism or their reproduction.

At the pre-social stage there is no question of a behavior failing or succeeding in relation to the demands of the *virtue* of consistency qua virtue. While some measure of consistency is necessary for any behavior, the virtue of consistency is more than just being able to behave. Judgments of failure or success in relation to the virtue of consistency are correctly directed to behavior or action once the kind of complex cognitive and social systems appear in which deviation and failure is an option.

⁵ For a detailed presentation of my view of the nature of emergentist explanation see Symons 2008; 2018a; 2018b

The virtue of consistency, I argue, can be explained in terms of its constitutive role in emergent forms of socially embedded agency and their associated identities. These identities are made possible by systems of social relations (institutions or social structures) that permit certain kinds of practices and make distinctive kinds of goods available. These institutions and structures appear over the course of natural and social history in ways that can be studied and explained scientifically. Once they are in place, the virtue of consistency becomes visible and important to us along with the relevant co-emergent institutions in ways I will describe here. This emergentist framework helps to address some traditional questions about the relationship between reason, the passions, motivation, and agency.

2. Trade-offs and consistency

Teachers of logic and critical thinking typically introduce inconsistency as a property of collections of sentences. Two sentences are inconsistent if they cannot both be true at the same time:

S_1 There are more cows in Ireland than in Greenland

S_2 There are not more cows in Ireland than in Greenland

Together S_2 and S_1 comprise an inconsistent set of sentences. Either of these sentences could turn out to be true under certain circumstances, but one knows that they cannot both be true together. More generally we say that a set of sentences is inconsistent if the sentences cannot possibly all be true at the same time.

Conversely, a set of sentences is consistent if they *can* possibly all be true at the same time. Being a consistent *person* means something quite different. To exhibit the virtue of consistency means to have a disposition to correct oneself in such a way that one tends to act in accordance with the epistemic principle that we ought to aim for consistency in our beliefs, utterances, preferences, etc.

The virtue of consistency can cohere with other epistemic virtues such that (for example) being consistent does not mean being stubborn or refusing to change one's mind in the face of strong countervailing evidence. Refusing to revise one's views over time is indicative of an epistemic vice rather than virtue. Thus, we can agree with Emerson that it is "foolish consistency" - meaning the virtue of consistency separated from the other epistemic virtues - that "is the hobgoblin of little minds". Refusing to change one's mind; refusing to modify one's beliefs in light of new evidence because of a desire to maintain the consistency of one's beliefs, would be an example of foolish consistency.

Notice that in the process of updating one's beliefs in a principled manner, say for example by following Bayes's rule in updating one's priors, the guiding principles of formal consistency will generally be assumed. There is no way to correctly update one's beliefs in accordance with Bayes's rule while violating the norm of consistency. Updating, changing one's mind, and learning more generally, requires some minimal

level of consistency. As we shall see, one follows the norm of consistency even in cases where one is deciding when to be inconsistent or when it is reasonable to tolerate inconsistency.

There are limits to our ability to ensure that we are following the norm of consistency because of the limits on our cognitive capacities. As a result, we must decide where to expend our resources. Let's begin by considering how one makes trade-offs with respect to consistency. A rational agent might decide that enforcing consistency in some domain would mean taking time and energy from other, more important pursuits. Consider how one might approach the question of whether it is worth checking to make sure all of one's past social media posts line up with one another, making sure that all the by-laws of all the student clubs in a university are consistent with some general principle, or organizing some elaborate stamp collection according to a preferred taxonomy. In each case one must rank the relative value of enforcing consistency against other uses of time and energy. This ranking might result in the decision to leave some inconsistencies be. This would be an example of a decision with respect to how one uses resources in relation to the norm of consistency. The goal of becoming consistent can be regarded as one good among others and it will be inevitably ranked according to one's hierarchy of values.⁶ Nevertheless, while becoming consistent is one good among many, the reasoning involved in judging the importance of consistency relative to other values requires that one abide by the norm of consistency. The rational agent adheres to the goal of consistency even if that agent must sometimes resign herself to leaving some domain or set of beliefs inconsistent because of other priorities or values. Deliberation and the rank ordering of values are governed by the need to be consistently in conformity with one's values. If one decides that organizing one's beliefs consistently in some domain would be too time consuming or expensive, one ought to do so rationally in order to honor one's commitments more broadly. A parent might be forced to renege on some promise to his children or might decide to treat them inconsistently because of an important concern that outranks the value of keeping his promise. The child's safety or well-being might be good grounds for his inconsistency for example. However, it should be noted that the parent would want his decision to act inconsistently in this case to have resulted from deliberations that were as internally consistent as possible. Thus, we must distinguish between being virtuously disposed towards consistency and acting in a manner that appears to conform to the virtue of consistency.

3. Why be consistent?

At this point, perhaps exceptional kinds of agents come to mind. We know people who do not seem to consistently order their preferences at all. They seem to make judgments in ad hoc ways, seemingly changing

⁶ One might prefer the approval of one's peer group to consistency, one might prefer to minimize one's effort and maximize leisure time above achieving consistency in some domain, or one might decide that resources are better spent on helping others or pursuing some other good end.

their minds about their commitments and adjusting their beliefs in unprincipled ways. What supports our intuitive judgment that this is not the right way to be? Let's consider some of the prudential reasons supporting a commitment to the norm of consistency. As we shall see, if we are unwilling to correct behavior that reveals inconsistency then in principle we will lose access to certain kinds of goods and modes of agency associated with some important social institutions. The simplest way of explaining this is in terms of a familiar notion from economics and decision theory: Inconsistent economic agents become targets for those who wish to turn those agents into money pumps. 'Money pump' is a term that is meant to describe the fate of someone who has some inconsistency in their preferences or some confusion about probability that leaves them vulnerable to exploitation (Cubitt and Sugden 2001).

What does it mean to have inconsistent preferences? Normally, if, one prefers strawberry ice cream to chocolate ice cream, one might pay a small sum to swap the chocolate for the strawberry. At the same time one would not pay to exchange strawberry for chocolate because this would run counter to one's preferences. Notice the relationship between the actions that one is willing to perform and the preferences that motivate those actions.

Now imagine someone with honestly held inconsistent preferences. In fact it is difficult to coherently imagine a typical adult having obviously inconsistent preferences such that they prefer strawberry ice cream to chocolate ice cream and they prefer chocolate ice cream to strawberry ice cream.⁷ Given such inconsistency, upon being asked whether they would like to pay a small amount to swap strawberry for chocolate they would accept the deal. When asked subsequently if they would like to pay a small amount to swap chocolate for strawberry they would also accept the deal. After these two payments are made they would return to their original position of having strawberry ice cream, but would be poorer as a result of the two exchanges. A competitor who discovered their inconsistency has converted them into a money pump and can drain them of resources until they correct their vulnerability in some way.

In such a simple case, it is difficult to imagine a typical adult human not quickly realizing what is happening and making a rapid adjustment. People are rarely in situations where publicly express explicitly contradictory desires and because of this economists and social scientists have no interest in obvious inconsistencies of this kind. Instead, social scientists are concerned with cases where the trouble is less obvious. Much more interesting is the set of cases where an agent's preferences are not transitive.⁸ Returning now to our preferences: Imagine that I am presented with some ice cream choices and I have some set of preferences with respect to flavors. Here's what I like:

⁷ Although parents can attest that such conflicting preferences can manifest themselves in the behavior of young children.

⁸ Transitivity is a property of some relations. For example, the relation of being *taller than* is transitive whereas the relation *being in love with* is not.

I prefer chocolate to orange

I prefer orange to banana

I prefer banana to chocolate

Let's first notice that my preferences are not transitive. If my preferences were transitive then when I prefer chocolate to orange and orange to banana, I should prefer chocolate to banana. My ranking of preferred flavors should be well-ordered if I am to avoid becoming a money pump. In our example, they are not. However, in some sufficiently long list of options one will inevitably lose track of the interactions of one's preferences. If I failed to notice this for some reason, then my preferences can be exploited. An agent with more cognitive resources than I have; an artificially intelligent agent for example, might be able to track and exploit past expressions of my preferences in ways that would leave me vulnerable to this kind of exploitation. In the simple case, if I stuck to my preferences and didn't notice what was happening to my financial situation, an unscrupulous ice cream dealer can use three flavors of ice cream to drain my bank account.

In dealings with other human beings I'm a savvy customer and can tell that there is financial trouble ahead for me if these shenanigans at the ice cream shop continue. I will therefore wisely refuse his offer of an exchange for chocolate. I can see where our transactions are leading – even though I *really do* prefer chocolate to orange. If you or I were to find ourselves in situations like this, we would quickly adjust *our behavior* in order to avoid becoming bankrupt. This is the corrective function of the virtue of consistency in action. If we adjust our behavior in this way, then it is because our desire to remain a viable player in the economic realm is strong enough to override our preference for chocolate ice cream over orange.

In economics, it is thought that some trade or series of trades in which one of the participants is always worse off than the other is very rare and should not be a persistent feature of markets. This is because one party will realize, as in the example above, that the inconsistent feature of one's preferences is incompatible with continued participation in market activities.⁹ It is worth noting that this need not be the case in the near future given the prospect of personalized attacks on individual consumer behavior by artificial intelligences armed with extensive access to one's prior behavior.¹⁰

It is a condition for participation in some basic social institutions, like markets, for example, that agents be reasonably conscientious about the consistency of their beliefs and preferences. Some are better

⁹ The phenomenon of intransitive preference is only one way in which irrational behavior can result in our being exploited. It is an instance of what is sometimes called a Dutch Book. A Dutch Book is a scenario in which, for example, the casino sets odds on a game of chance such that the set-up guarantees a profit for the house, regardless of the outcome of the gamble. In order to be turned into a money pump you need only to be irrational about the way that probabilities work and to be unlucky enough to have that irrationality exploited (Hájek 2009).

¹⁰ A demonstration of the feasibility of such approaches is given in Dezfouli et.al (2020)

than others, but complete lack of conscientiousness with respect to consistency means being selected out of the market. Notice that by accounting for it in this way, the virtue of consistency is constitutive of market agency. Caring about consistency is a condition for the possibility of enduring participation in those organized interpersonal contexts within which one can be an agent capable of expressing interests and preferences.

Markets and other systems of social cooperation can help us to recognize that we have held false beliefs, held inconsistent preferences, or have otherwise reasoned badly in the past. The money pump is a simple way of illustrating the role of the epistemic virtue of consistency in a social context. On one level we might be tempted to say that the tendency towards consistency in agents is a formal feature of certain kinds of emergent social equilibria. In the case of market behaviors, for example, having intransitive preferences will count as an exploitable failure insofar as successful, or at least persistent, market agency requires respect for consistency. At the same time, agents are related to these emergent social systems in ways that allow them to succeed or fail. Agents will sometimes fail to respond to market signals by exhibiting the required corrective tendencies associated with the virtue of consistency within the appropriate timeframe.

Unlike the Kantian picture of the constitutive relationship between rational principles and willing which sees such failures as impossible for properly rational agents, failure is a common feature of market agency. The Kantian, would say that a will that is not governed by rational principles is not a will. However, the fact that someone happens to have intransitive preferences certainly does not disqualify them from being market agents. One does not immediately cease to be a market agent in virtue of having an irrational will. Arguably, all non-logically omniscient market agents who have a reasonably complex set of preferences will have some kind of intransitivity that might be revealed to them by a suitably complex and dynamic market. However, one will not be an agent for very long if one does not correct one's inconsistency when it is revealed. The personal virtue of consistency in the market context is the disposition to organize one's market agency in such a way as to correct inconsistency or to *appear to be* as consistent as possible. Part of being a market agent is that one adjusts in light of what market interactions teach us.

Given that market interactions have a competitive component, the price of inconsistency is that one can suffer being selected out of the market through immiseration. Thus, consistency is not constitutive of specific *acts* of market agency or willing; it is not constitutive of the market agent's will in the Kantian sense. Instead, the market agent strives for consistency in order to remain a viable long-term actor in the market. Successfully striving for consistency is largely a matter of correcting behaviors that manifest inconsistency. The virtue of consistency, in this sense, can be said to be constitutive of *long-term* market agency.

For Kantians like Christine Korsgaard, for example, rationality is constitutive of the normatively ideal will. She recognizes, of course that we fail to live up to this normative ideal most of the time. However, given that we are considering the will in terms of normative ideals, the question of why one ought to will in accordance with rationality has no real content because to will simply is to will in accordance with principles

of rationality. The Kantian argument from the constitution of the ideal will assumes that there is no willing that is not constituted by the principles of rationality. What happens, in the case of messy ordinary decision-making is, according to the Kantian, simply not the kind of willing that we should be concerned with.¹¹ If one accepts this position, then explanations of the virtue of consistency of the kind that I have provided here are not on the table. It strikes me as a failure of a philosophical view that it rules out possible lines of explanation when such explanations can be given. It is an even more serious failure if it makes obviously true judgments false. It is obviously true that one can will in ways that are not in accordance with principles of rationality. This is one of the reasons we praise people who exemplify the virtue of consistency. The virtue of consistency is simply not a virtue one can ascribe to willing insofar as inconsistent willing is not willing.

One reason that the Kantian formulation of questions around the will, rationality, and action seems wrong to so many readers is that it presents an image of decision making for finite agents that is completely alien to ordinary human experience. I find that my own actions are regularly out of alignment with principles of rationality. This is because I am not logically omniscient, I have an unreliable memory, I am subject to cognitive biases of various kinds, and I have a wide range of preferences that come into conflict in practice etc. Insofar as we are embodied beings with finite capacities, Kantian rigorism certainly seems wrong. By contrast an emergentist story about social institutions like markets explains why avoiding behaviors that manifest inconsistency is associated with long-term access to market agency and the goods that come along with the relevant kinds of associated identities.

The Kantian account of motivation and action is frequently contrasted with a broadly Humean view that explains our motivation via the passions. Rather than insisting on the regulative role of the principles of rationality, the Humean argues for the primacy of the passions in human action. On this view, any putative story about consistency, rationality, and agency can ultimately be traced back to non-rational preferences and passions. Putting the contrast very crudely, we can say that the Humean rejects the Kantian account of the connection between reasons, principles of rationality, and the will, and instead, seeks explanation in the activity of the passions.

Unlike the Humean picture of the will, the view I have described recognizes that formal constraints in the context of market agency give reasons to be consistent that are internal to specific kinds of practices and institutions at play. In the example we considered, we saw how institutions like markets are constituted in ways that require agents to attend to epistemic norms in order for them to persist as *market* agents. Given this constitutive aspect of epistemic normativity in social agency, the view has a Kantian flavor. But, as

¹¹ I am grateful to Brad Cokelet for pointing out that Korsgaard's position on the constitutive role of the rational principles for the will should not be understood as committing her to any position with respect to the mechanics of ordinary decision-making. While there are sharp differences between the Kantian and the emergentist position on the nature of the will, it would be a mistake to criticize the Kantian project as failing on empirical grounds. I do not mean to offer a straw man form of Kantianism here. Instead, we should recognize that Kantian arguments are operating at an idealized normative level that have little in common with the kind of empirical approach that I am engaged with here.

mentioned above, it is distinctively non-Kantian insofar as it allows for the human will to frequently fail to be rational. There are reasons for being consistent that are neither broadly Kantian, nor Humean. The contexts within which these reasons figure are not part of a pre-social nature, rather they emerge over the course of natural and social history, but they are not merely socially constructed. They are novel relative to what came before but they are not mysterious. Rather, they admit of natural explanations in the sense that I have tried to describe elsewhere (Symons 2018).

4. Explaining rationality and explaining consistency

Why do we care so much about being rational and following associated norms, like consistency? As mentioned above, there are more and less fruitful ways of approaching this problem. To begin with, we can ask questions of this kind in ways that are simply not well-formed and do not help us to arrive at explanations. When we considered the Kantian account of the constitution of the will in accordance with rational principles we also noted the ways in which it closes off the kinds of explanations that we explored in the case of market agency.

Niko Kolodny, for example, has argued along similar lines suggesting that the traditional way of framing the “Why be rational?” question does not admit of a coherent response. On his view, there simply cannot be a reason to be rational. The way the world is can give us reasons against one part of the conjunction A and *not* A , for example, but there is no reason, according to Kolodny for an agent to be rational (2005). Let’s assume, for the sake of this paper, that Kolodny is sound and that we cannot give reasons to be rational. Can we provide legitimate reasons for valuing and being motivated by consistency? The prospects look brighter, as we have seen above. Questions about rationality can be addressed once we formulate them in ways that do not devolve into self-undermining requests for reasons in support of giving reasons. By exploring reasons for being consistent, we have shed some light on the broader question of what it means to be rational and why it matters.

The first step is to recognize that consistency and rationality are distinguishable. A rational agent can aspire to the virtue of consistency for good reasons. Commonsense reasons to value consistency come immediately to mind, including the connection between consistency and our desire to accomplish our goals. We recognize that inconsistency can hamper decision-making and action. One cannot both drink one’s beer and not drink one’s beer at the same time without risking a spill. Thus, given our interest in decisions that inform our actions, we will be motivated to cultivate the virtue of consistency. In general, there are good practical reasons to strive for consistency in the pursuit of one’s ends. Relatedly, there are also moral reasons to avoid inconsistency. It would be irresponsible, and therefore potentially unethical, to fail to pursue consistency in decision-making in some contexts. It is obviously blameworthy, for example, to be negligent

in matters that have serious consequences for the well-being of others. In this respect, commonsense suggests that there are often moral reasons grounding the epistemic virtue of consistency.

Returning to Kolodny's argument against reasons to be rational, we might wonder whether the commonsense reasons to be consistent can also serve as reasons to be rational in Kolodny's sense. It turns out, they don't. Let's consider them one at a time:

1. We should be rational so that we can act in ways that conform with our intentions
2. We should be rational because irrationality can lead us to act unethically.

The trouble with (1) is that it assumes what it is known as the instrumental principle, that one should always will the means one knows are necessary for satisfaction of one's ends.¹² The lack of a satisfactory account of why rationality commits us to the instrumental principle has been a significant feature of the recent debate. What we saw above is that the virtue of consistency is useful in the pursuit of one's practical goals, this assumes, but does not explain the instrumental principle. Once we have the instrumental principle in place, we can explain the virtue of consistency by reference to that principle. However, consistency is not enough to connect intending an end with intending the necessary means.

Kolodny would correctly argue that the commonsense reasons for consistency given so far are not responding to the demand for distinctive reasons to be rational. Let's briefly explore what one might mean by *distinctive reasons to be rational*, before considering moral reasons for consistency. Kolodny pointed to two ways of asking questions about reasons. He first asks us to consider the following ordinary kind of practice of giving reasons:

Why should we leave the building? Because there's a gas leak.

Here, the gas leak counts as a reason to leave. Now consider another kind of question about reasons:

OK, you want to leave, so open the door? Why should I open the door? Because if you do not open the door, you will not be able to leave? Since you want to leave, and since the only means available to you is to open the door, you must open the door? It would be irrational to do otherwise

¹² This principle has been the subject of considerable controversy recently. Jonathon Way explains the central problem for this assumption when he describes how "[t]he so called Wide-Scope view of instrumental reason holds that you should not intend an end without also intending what you believe to be the necessary means. The trouble for Wide-Scopers is that they have been unable to meet a simple Explanatory Challenge: why shouldn't you intend an end without intending the necessary means? What reason is there not to do so?"

Kolodny argues that any reason we might give for being rational, for example in cases like the second case reduces to reasons of the first kind, namely reasons mentioning states of affairs about the world. At this stage we have two related challenges to taking 1 as a reason to be rational and not merely consistent: The explanatory challenge related to the instrumental principle and the putative reducibility of talk about irrationality to reasons mentioning states of affairs about the world.

Without assessing the strength of these arguments at this point, let's move on to challenges to the second reason for consistency 2 (recall that 2 offers moral grounds for the epistemic norm). The objection here is that in doing so, we will assume the very epistemic principles we are trying to ground. In its general form the circularity objection is that moral reasoning cannot ground epistemic norms since moral reasoning only works if it is in compliance with epistemic norms. This seems excessive in the case of the epistemic virtue of consistency. Surely, there is nothing viciously circular about saying that in some cases it is morally obligatory to aim to deliberate in a consistent manner. The problem of saying that some set of moral norms *G* grounds the epistemic norm *C* in such a way that *G* does not assume *C* in some way. Acting on *G* would require rationality, or at the very least would require an agent to understand *G*. As we shall see, the apparent circularity of the moral grounding strategy can be resolved relatively straightforwardly. However, we might still wonder whether a moral principle can be sufficient to ground the norm of consistency. In fact, I think that both the instrumental principle and the moral grounding arguments are good commonsense reasons to be consistent. However, we can be moved at a commonsense level by 1 and 2 while recognizing that they will not count as convincing philosophical reasons to persuade our opponent that they ought to be rational.

5. Emergentist Explanations

One way that our opponent can respond to the commonsense reasons to be consistent is to follow Kolodny in arguing that insofar as 1 and 2 are persuasive, they persuade in ways that have nothing to do with consistency itself. They rely, for example, on reasons having to do with facts about the world or perhaps with reasons to be good. In addition to the kinds of objections that Kolodny raises, there are more traditional Kantian reasons to reject the *why be rational?* question as we saw above. As traditional Kantian arguments and Kolodny's argument make clear, the kinds of commonsense reasons to be consistent that we have been considering only make sense when given by an agent who is already embedded within a world where the norms of consistency are already in place. Emergent social institutions make their associated kinds of agency possible and the epistemic virtue of consistency is constitutive of that agency in ways we described above. The kind of market agency discussed in Section Two exemplified the role of the virtue of consistency in the long-run maintenance of an emergent form of agency. In this sense, the virtue of consistency is constitutive of that agency.

This paper has approached consistency very differently from the way that most virtue epistemologists understand their task. For the most part, they adopt what can very loosely be called an Aristotelian explanatory strategy. The emergentist strategy presented above contrasts sharply, for example, with Foot's advice explanations of the virtues. The virtues, on her view, are correctives to weaknesses or defects in our natural condition; courage is a virtue because we have a natural tendency to behave in cowardly ways. By contrast, in the account we have given here, the corrective function of the virtue of consistency only arises in the co-emergent social institution. We cannot find a natural paradigm of the virtue of consistency, nor is it obvious that there is a countervailing kind of defect that thwarts us in our quest for consistency. Granted, we are naturally limited in our cognitive capacity in ways that make it hard to be both sufficiently complicated for adult life and consistent, but what is natural here is our finitude, not our inconsistency. However, as I have argued above, we can be judged inconsistent or virtuously consistent only in relation to the social institutions that are home to these norms. In this sense, we cannot look to pre-social nature for orientation.

Throughout her work, Foot looks to the pre-social and pre-conventional world for guidance as we attempt to formulate an account of the kinds of goodness we associate with the virtuous life. Her goal was to show moral evil as a "natural defect" (2001, 5) and that one's failures are at least in part failures of one's capacity to successfully reason about practical matters. To fail here is to fail functionally in relation to the norms we associate with the species in the same way that we would identify a biological defect in relation to the norms associated with its kind. The view I present here does not ground the virtue of consistency in some natural capacity. Instead, for the emergentist, part of our species identity involves the open-ended development of new social systems and institutions with their associated norms and identities. The kinds of excellences or virtues that human beings can aspire to are not fixed by some stable species identity. In this sense it is a mistake look to plants and animals for guidance in human affairs.

Acknowledgments:

I am very grateful to audiences at Houston Community College, University of Kansas, and Universitatea de Stat din Moldova. Thanks especially to Hassan Alsharif, Brad Cokelet, Dale Dorsey, Dan Flores, Erin Frykholm, Cliff Hill, Jason Raibley, and Irina Symons, for critical comments on earlier drafts of this paper.

Bibliography:

Alsharif, H., & Symons, J. (2021). Open-mindedness as a Corrective Virtue. *Philosophy*, 96(1), 73-97.

Calvo, P., Martín, E., & Symons, J. (2014). The Emergence of Systematicity in Minimally Cognitive Agents, in *The Architecture of Cognition: Rethinking Fodor and Pylyshyn's Systematicity Challenge*, P. Calvo and J. Symons (eds.), Cambridge: MIT Press. 397-434.

- Cubitt, R. P., & Sugden, R. (2001). On money pumps. *Games and Economic Behavior*, 37(1), 121-160.
- Dezfouli, A., Nock, R., Dayan, P. (2020) Adversarial vulnerabilities of human decision-making. *Proceedings of the National Academy of Sciences* 117 (46) 29221-29228; DOI: 10.1073/pnas.2016921117
- Foot, P. (2001). *Natural goodness*. New York and Oxford: Clarendon Press.
- Foot, P. (2002) *Virtues and vices and other essays in moral philosophy*. New York and Oxford: Oxford University Press.
- Hájek, A. (2009). Dutch Book Arguments, in *The Oxford Handbook of Rational and Social Choice*, P. Anand, P. Pattanaik, and C. Puppe (eds.), Oxford: Oxford University Press, 173–195.
- Horgan, T., Potrc, M., & Strahovnik, V. (2018). Core and ancillary epistemic virtues. *Acta Analytica*, 33(3), 295-309.
- Kolodny, Niko. (2005) "Why be rational?." *Mind* 114, no. 455 509-563.
- O'Neill, O. (1985). Consistency in action, in *Morality and Universality*, Theory and Decision Library (An International Series in the Philosophy and Methodology of the Social and Behavioral Sciences), vol 45. N.T Potter and M. Timmons (eds.), Dordrecht: Springer, 159-186 https://doi.org/10.1007/978-94-009-5285-0_8
- Rendsvig, R. and Symons, J. (2019) "Epistemic Logic", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2019/entries/logic-epistemic/>>.
- Symons, J. (2008). Computational models of emergent properties. *Minds and Machines*, 18(4), 475-491.
- Symons, J. (2018a). Brute facts about emergence. In E. Vintiadis & C. Mekios (Eds.), *Brute facts*. Oxford: Oxford University Press, 177-195
- Symons, J. (2018b) Metaphysical and scientific approaches to emergence: Varieties of fundamentality and theoretical completeness in *Emergent Behavior in Complex Systems Engineering: A Modeling and Simulations Approach*. S. Mittal, S. Diallo and A. Tolk (eds.) Wiley, 3-20. DOI: 10.1002/9781119378952.ch1
- Turri, John, Mark Alfano, and John Greco, (2019) "Virtue Epistemology", *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2019/entries/epistemology-virtue/>>.
- Way, Jonathan (2012) Explaining the Instrumental Principle, *Australasian Journal of Philosophy*, 90:3, 487-506, DOI: [10.1080/00048402.2011.598174](https://doi.org/10.1080/00048402.2011.598174)