

SPECULATIVE MISTAKES AND ORDINARY TEMPTATIONS:

KANT ON INSTRUMENTALIST CONCEPTIONS OF PRACTICAL REASON

Sergio Tenenbaum

Instrumentalism, roughly, claims that we are rational as long as we use the best means to get what we want, or what makes us happy or satisfied, without specifying that it's intrinsically rational or irrational to want any specific thing. Thus for the instrumentalist the only legitimate principle of rationality is some version of the principle of instrumental reasoning (PIR): one should not pursue an end without also pursuing adequate means to this end. Since it seems obvious that our lives go better when we get what we want, or when we are happy (and worse otherwise), it isn't hard to see the appeal of instrumentalism and the difficulty of arguing against it. Those who consider this appeal deceptive have often turned to Kant's work as a prime source of anti-instrumentalism arguments. This paper tries to examine the nature of Kant's case against instrumentalism; in particular it argues that, contrary to what has been recently argued, Kant thinks that instrumentalism is a *coherent* (albeit false) position.¹

Kant's main target is better characterized not as instrumentalism, but as empiricism: a view according to which all incentives of practical reason are empirical.² The first section tries to understand more precisely what Kant finds problematic about empiricism. It starts by examining an apparent tension in Kant's criticisms of empiricist philosophers. Kant seems to think that the mistakes that empiricist philosophers make are *never* made by ordinary reason, yet Kant describes the temptations of

ordinary reason in a way that makes them seem indistinguishable from the mistakes of the empiricist philosophers. Solving this tension sheds light on Kant's rejection of empiricism. Moreover two things should emerge that will help us understand what kinds of Kantian arguments against instrumentalism are and are not possible. First, Kant is committed to attributing different conceptions of happiness to the empiricist philosophers and to ordinary reason. Second, Kant is committed to the *possibility* of a purely instrumental practical reason. This second point should reduce the hopes for finding in Kant an argument for the incoherence of instrumentalism, and Kant has good reasons to avoid this line of argument. However the first point might provide materials for a more promising line of attack on instrumentalism. For if it's right, instrumentalism cannot take as its point of departure the obvious fact that our happiness as conceived by ordinary reason matters to us. Moreover instrumentalism is a live option only insofar as we can make sense of a notion of a given end. Although Kant took this notion to be unproblematic, it's far from clear that contemporary instrumentalists have it readily available.

AN APPARENT TENSION

Kant discusses extensively the "Principle of Self-Love" (PH)³—the principle that makes our own happiness the determining ground of the power of choice. When he criticizes philosophers for taking principles falling under this heading to be practical laws, Kant always has kind words for ordinary reason—our pre-philosophical exercise of our rational capacities. Commenting on the conflict between the moral law and PH, Kant claims that:

[this conflict] is practical and would ruin morality altogether were not the voice of reason . . . so audible even to the most common human beings; thus *it can maintain itself only in the perplexing speculations* of the schools, which are brazen enough to shut their ears to that heavenly voice.⁴

Kant even asks whether it would not be better if philosophers were to leave ordinary reason to its own devices in matters of morality. However, Kant *doesn't* think that the question should be answered in the affirmative. Ordinary reason has its own temptations to which, if left alone, it would likely succumb:

There is something splendid about innocence; but what is bad about it . . . is that it cannot protect itself very well and is easily seduced. The human being feels within himself a powerful

counterweight to all the commands of duty . . . the counterweight of his needs and inclinations.⁵

These “apparently plausible” claims of the inclinations give rise to a “natural dialectic”:

a propensity to rationalize against those strict laws of duty and to cast doubt upon their validity, . . . to make them better suited to our wishes and inclinations.⁶

It seems that the problem with ordinary reason—what makes it need help from philosophy—is that in its pre-philosophical innocence it has a tendency to put the claims of the inclinations, at least some times, above the claims of the moral law. Unaided by philosophy, we might give in, for instance, to the thought that, given the way the world is, a comfortable life requires a few exceptions to a general policy of honesty. But here we might begin to lose our grip on why ordinary reason is doing any better than the “perplexing speculations of the schools.” It seems that giving in to this temptation would amount to making the principle of one’s happiness into a practical law, or at least into the supreme principle of practical reason. After all, for Kant, happiness just is “the entire satisfaction of our inclinations.” To make the commands of duty better suited to our “wishes and inclinations” is to give preference to our own happiness when paying heed to the moral law proves to be burdensome, or, in other words, to subordinate the moral law to our own happiness.

But if this is the case, why is the fate of pre-philosophical ordinary reason very different from that of the doctrine of the philosophers? Why isn’t it the case that both have a tendency to succumb to the same temptation: to take PH to be the supreme principle of practical reason? There seems to be no reason to extol the accomplishments of ordinary reason while remonstrating fellow philosophers.

One might argue that there is here a confusion between happiness and PH. Philosophers tend to give an account of morality in terms of PH, but ordinary reason is tempted by *happiness*, not by a philosophical doctrine. This suggestion isn’t far from the truth, but it can’t be the whole story. For Kant, the temptation that happiness poses us cannot be detached from the adoption of a certain principle. In later works, Kant talks about “self-conceit” or a “propensity to evil” in characterizing a tendency in human nature to stray from the moral law: this propensity is a propensity to *make self-love the condition of the moral law*; that is, to make self-love the supreme practical

principle. But this seems to describe precisely PH insofar as it characterizes a philosophical mistake; for the mistake is to have “the principle of one’s happiness made the determining ground of the will.”⁷

THAT WHICH WE ALL KNOW

In order to generate the tension described in the preceding section, one must assume that the empiricist notion and our ordinary notion of happiness are the same. The present aim is to show that, on Kant’s view, empiricist philosophy generates a notion of happiness that is importantly different from how we ordinarily conceive of happiness. In order to understand this difference, one needs to examine what it is Kant thinks that we all know, but that the empiricist misses when philosophizing. This seems to be precisely what Kant is describing in the following passage:

Suppose someone recommends to you as steward a man to whom you could blindly trust all your affairs and, in order to inspire you with confidence extols him as a prudent human being with masterly understanding of his own advantage. . . . You would believe that the recommender was making a fool of you.⁸

Kant goes on to conclude as follows:

So distinctly and sharply drawn are the boundaries of morality and self-love that even the most ordinary eye cannot fail to distinguish whether something belongs to the one or the other.⁹

Here we can form a conjecture about the nature of our ordinary wisdom: What we are being credited with is an ability to *distinguish* between the two incentives. The central mistake of the philosopher who takes the principle of happiness to be the sole practical principle is to try to turn the incentives of morality and of self-interest into two *homogeneous* incentives, as being of “one and the same kind.”¹⁰ On the other hand, Kant describes virtue and happiness as two “extremely heterogeneous concepts.” The reasons for claiming the homogeneity of the incentives of self-love are complex, and cannot be discussed in detail here. In a nutshell, according to Kant, insofar as we act on the grounds of these incentives we act on the same grounds. When one chooses to pursue the object of an inclination on the basis of self-love, one chooses on the grounds that satisfaction of inclinations contributes to my happiness. Although this claim doesn’t imply that we always set the same end when we act from self-love,¹¹ it does imply that one *can* always compare any two incentives of self-love

insofar as they serve as grounds of determination of the will. Since choice in terms of these incentives is choice in terms of their contribution to our happiness, we can always ask which one will make a greater contribution to our happiness. This claim also implies that there is no content-based way of classifying desires into higher and lower ones so that satisfying desires of the former kind is or should be *necessarily* preferred to satisfying desires of the latter kind. Let's take, for instance, a putative classification of incentives of self-love with which Kant was particularly concerned: the classification between higher and lower pleasures. Higher pleasures are supposed to be intellectual pleasures, pleasures which human beings are capable of enjoying and animals not, or even pleasures of "social" or "moral" behavior, pleasures from which the unsocial or immoral person is shut out. These pleasures are supposed to be preferable, *qua pleasures*, to the counterpart "lower" ones. Although we are familiar with this distinction from Mill, Kant's predecessors made rather similar points. This is a passage from Shaftesbury's "Inquiry Concerning Virtue and Merit":

How much the social pleasures are superior to any other may be known by visible tokens and effects. The very outward features . . . are expressive of a more intense . . . pleasure than those which attend the satisfaction of thirst, hunger and other ardent appetites . . . *No joy, merely of sense, can be a match for it.* Whoever is judge of both the pleasures will ever give the preference to the former. But to be able to judge of both, it is necessary to have a sense of each.¹²

The claim that no sensual joy *can* match a social pleasure introduces the kind of heterogeneity that Kant doesn't allow for. If the claim is about a *necessary* superiority of the social pleasure it misses the nature of empirical grounds of determination. Since empirical grounds of determination depend on how certain objects happen to affect our faculty of desire, there seems to be no grounds to establish that certain objects *must* affect our faculty of desires favorably or unfavorably; we can at most through experience find out that they tend to affect our faculty of desires in a certain way. If the claim is supposed to make an empirical generalization, it's just false. Kant gives the following exceptions to the generalization:

The same human being can return unread an instructive book that he cannot again obtain in order not to miss a hunt . . . ; he can even repulse a poor man whom at other times it is a joy

for him to benefit because he now has only enough money in his pocket to pay for his admission to the theater.¹³

Note that Kant doesn't choose the examples of rakes, villains or even of actions performed by average humans but which bring shame to them. Some more high-minded persons might take issue with the actions described in this quote, but for most of us, it's not even the case that our second order affections condemn those first order affections. Thus, in the relevant sense of "reflective" for Shaftesbury, these are affections that receive our reflective approval.

Heterogeneous determining grounds of the will cannot be compared in this manner; they can only stand in relations of subordination and superordination. Since choice based on the grounds cannot be ultimately reduced to a common ground, they can never be integrated in a form of deliberation that assigns relative weights to different incentives. And, indeed, this is how Kant describes the adoption of an evil disposition in the *Religion*:

He indeed incorporates the moral law into those maxims, together with the maxim of self-love; since, however, *he realizes the two cannot be juxtaposed* but one must be subordinated to the other as its supreme condition, he makes the incentives of self-love . . . the condition of compliance with the moral law.¹⁴

Given that the moral law is primarily a constraint on the *form* of the will, a constraint on whether our maxims have the form of a universal law, it cannot be compared to the material grounds of determination of the will, grounds based on a particular content of the faculty of desire such as the object of an inclination. Doubtless the moral incentive is grounded on an unconditional principle, and thus must take precedence over any incentive that is merely *given* to reason; any other incentive ought to be incorporated into one's maxim only insofar as the moral law allows it. However there could be no "coalition" of motives—no way of assigning certain relative weights to these incentives short of giving full priority to one of them. Given that self-love and the moral law are the only two determining grounds of the will, there could be no third principle that would determine the proper "weight" of each motive.

One might think that Kant's contention that there could be no coalition of motives leaves us with the following two options: either one is good and pursues one's happiness only when there is no more duty left to be fulfilled, or one is evil and uses the moral law only as a tie-breaking principle for actions to which

one is indifferent. Fortunately, we do not need to attribute this view to Kant. Our conception of our happiness is shaped by the fact that we are guided by a moral incentive—it will be shaped by our incentive to give to our happiness the form of universality. On the other hand, our particular duties cannot be determined solely through inspection of the moral law. For instance, friendship, for Kant, is a moral ideal, because it manifests the principle of benevolence, the principle that requires that we make other persons' ends our own. But that friendship is a moral ideal is in part determined by the shape of our sensibility and by our needs—in particular that we can expect to be satisfied with some forms of attachment but not others.¹⁶ However, no matter how much the two incentives are shaped by each other, given that we're not holy creatures, we know that the two incentives are not in perfect harmony. Reason requires that we subordinate our happiness to the moral law and sacrifice our happiness in cases of conflict. However, as imperfect creatures, we might end up inverting the order of incentives and making exceptions to the universal character of the moral law on behalf of self-love.

But the empiricist doesn't rely on any such coalition of motives. She doesn't claim that self-love and morality are equally valid incentives, and of course she also doesn't claim, in any straightforward sense, that our happiness is more choiceworthy than a moral life. She claims that self-love is the *only* incentive, and our happiness our *only* end. If Kant is right, this cannot be an accurate account of *human* practical reason, since it ignores our awareness of the moral law. However, this aspect of the empiricist view ensures that it doesn't conceive of our happiness as unconditionally good—as something that reason itself sets as an object of the faculty of desire. For the empiricist everything that is good is merely conditionally good; it's good for something that we find ourselves pursuing. Our ultimate ends aren't set by reason, but merely *given*. This is so far a coherent position; indeed this is how one ought to conceive of happiness as an end unconstrained by the moral law.

Empiricists typically also want to distinguish between “lower” and “higher” forms of happiness. This is, by Kant's lights, not surprising; the empiricist wants *somehow* to preserve the ordinary wisdom of which she herself undoubtedly partakes, and thus she will attempt to carve out some room for the superiority of the moral incentive within empiricist bounds. It's *this* move that leads her to an inconsistent position; as we saw, all the incentives of an

empirically conditioned practical reason are homogeneous. The empiricists described here are instrumentalists. Reason in their view doesn't set ends; our ends are simply given to us by our nature. It's not instrumentalism that leads the empiricist to inconsistency, but the attempt to find room within this view for the subordination of all ends to the ends of morality.

ORDINARY FAILINGS

The empiricist commitment to the homogeneity of all incentives isn't shared by ordinary reason. We always recognize the difference between morality and self-love. So how could ordinary reason go wrong? Kant claims that our most fundamental disposition must contain both the incentives of morality and self-love.¹⁶ Since we are always aware of the moral law, we cannot ignore its claims completely, and since we are finite beings, we cannot leave aside the demands of self-love. Thus the difference between a good and an evil disposition cannot rest on the incentives they include, but on how these incentives are ordered. Since we already ruled out above the possibility that these heterogeneous incentives be ordered in any relation other than a relation of subordination, an evil disposition must be a disposition in which one claims priority to our happiness over the demands of the moral law. Insofar as reason represents our happiness as something that we should pursue at the expense of the moral law, it must conceive it as making a more legitimate claim. This involves at least representing our happiness as good, as having the same kind of rational warrant as the moral law. But how could the claims of happiness have this kind of rational warrant? According to Kant, the only coherent way in which we can represent our happiness to be good is to subordinate the claims of happiness to the claims of morality, by providing the matter of the faculty of desire with the universality of form that pure practical reason demands.¹⁷ This is indeed how, in the *Critique of Practical Reason*, Kant characterizes the derivation of the principle of beneficence:

Let the matter be, for example, my own happiness. Thus if I attribute it to each (as in the case of finite beings, I may in fact do), it can become an *objective* practical law if I include in it the happiness of others.¹⁸

This is indeed a conception of happiness in which one's happiness is good, a source of conditioned but legitimate claims. But it's, of course, not a temptation; it's how the virtuous person

conceives of her happiness. This conception of happiness is different from the one put forth by the empiricist philosopher: it's a conception of a good as opposed to a conception of the "given" of practical reason. The ends set by the virtuous person, including those that constitute her happiness, are ends that she correctly considers to be good. These are ends to which even an impartial, rational being should not be indifferent,¹⁹ and thus they are necessary objects of the will of every rational being, and consequently good.²⁰ In order to form a coherent conception of our happiness as good, we must accept the subordination of our happiness to the moral law. Thus the person who succumbs to temptation cannot form a coherent conception of the claims that her happiness makes on her will, for she must take her happiness to be good independently of the moral law—thus she takes it to be *unconditionally* good. Indeed Kant's discussion of self-conceit describes it as a propensity to usurp to self-love the claims that only morality can make upon us. Self-conceit makes self-love "itself lawgiving and the unconditional practical principle."²¹ It's no surprise that Kant regards evil as always involving a form of self-deception since the agent always implicitly knows that only the object of the moral law is unconditionally good:

Any profession of reverence for the moral law which in its maxim doesn't however grant to the law . . . preponderance over all other determining grounds of the power of choice is hypocritical, and the propensity to it is inward deceit, i.e., a propensity to lie to oneself in the interpretation of the moral law, to its prejudice.²²

Because we cannot extirpate morality from its soul, and thus cannot blind ourselves to the heterogeneity of these two kinds of demands, when we succumb to temptation we must grant self-love the rational status that only the moral law can have. Thus, even what Kant would call "a perverse power of choice" cannot conceive of its happiness in ways described by empiricist philosophers. It cannot, in particular, see happiness as merely a given end which reason has the commission to promote.²³ For this would represent the claims of happiness as inferior to the demands of the moral law, which presents its object as absolutely good—not as an external optional commission, but as an internal requirement of practical reason. Since we cannot forget our awareness of the moral law, happiness must appear to ordinary reason as good in the same way if one can "rationalize against those strict laws of duty and cast doubt upon their validity."²⁴ Thus the vicious person shares with the virtuous person,

but not with the empiricist, a conception of her happiness as good. On this account, the vicious person does adopt an incoherent stance; one's happiness can be good only if constrained by the moral law. But what makes the vicious person's stance incoherent is exactly the feature of her view that doesn't figure in the empiricist view.

KANT AND THE MODERN EMPIRICISTS

Kantians tend to try to make use of Kant to show that instrumentalism is, in some sense, incoherent.²⁵ However, if instrumentalists are the intellectual descendants of Kant's empiricist contemporaries, we should be suspicious of such attempts. For our discussion would show that the only advantage that empiricism could possibly claim over the exercise of ordinary reason is exactly coherence. Kant did think that empiricists would often end up holding views that were in fact incoherent, and he notes that examples of consistent systems are more often found among the ancients than among his contemporaries.²⁶ However, the main source of incoherence that Kant finds in empiricism is its attempt to ground the subordination of the claims of self-love to the claims of morality within the framework of the principle of happiness. The attempt to find a higher place for morality within homogeneous sources of motivation is what, according to Kant, leads those philosophers to problems. However, our contemporaries often make no such attempts.²⁷ Even those contemporary empiricists who would like to approximate the rationalist picture in which morality occupies such a prominent place might find distinctions of higher and lower pleasures rather outmoded.

Given that Kant fires so many shots against PH, it's hard to resist the temptation to believe that some must be capable of hitting contemporary targets. A seemingly promising way to use Kant's ammunition, acutely presented by Christine Korsgaard, is to propose a reflective question to the end presented by the empiricist;²⁸ we ask about the proposed end whether we indeed have a reason to pursue it. But this kind of argument doesn't appear to be in Kant's *corpus*. It will be useful at this point to discuss in more detail how and why Kant thinks that a being whose practical reason is purely instrumental is possible. This will help us show, first, why Korsgaard cannot close out the possibility of such a being. Moreover this discussion will also suggest a more promising line of criticism of contemporary instrumentalism.

Let's start with a passage that seems to present an argument against instrumentalism.³⁰ After some conjectures on the first occasion in human history when "reason has played tricks on the voice of nature" and man chose not to act according to instinct, Kant says:

He discovered in himself an ability to choose his own way of life and thus not to be bound like other animals to only a single one. . . . He stood as if at the edge of an abyss; for besides the particular objects of desire on which instinct had until now made him dependent, there opened up to him an infinitude of them, among which he could not choose, for he had no knowledge whatsoever to base choice on; and it was now equally impossible for him to run back from his once tasted state of freedom to his former servitude (to the rule of instincts).³⁰

Indeed the temptation to read into it a criticism of the coherence of instrumentalism seems overwhelming: Kant describes what is tasted here as "a state of freedom," as opposed to the pre-existing servitude. It seems easy to conclude that only the moral law, the law of freedom, can solve the problem raised by the discovery of his ability to choose. Thus it would seem that instrumentalism cannot answer a question raised by the most minimal employment of practical reason—by the smallest deviation from blind instinct. However, this passage also resists being read as displaying any such problem. It's first important to note that prudential reasoning also has to question instinct and make choices under a vague conception of an ideal of happiness. The being who forms a conception of this ideal through the merely regulative employment of reason no longer follows the "rule of instinct," and there is no reason to think that the lack of "knowledge" alluded to in the above passage isn't just the lack of empirical knowledge of a being who has been blindly following an instinct and now must figure out what would make her happy on her own. The continuation of the argument of "Conjectural Beginnings" seems to be compatible only with this more modest reading of the passage. For Kant goes on to describe a progression of reason's employment in which the moral law stands only at the far end of the path. Even its first glimmerings appear only at the next stage of this progression.³¹

Given that human beings are free agents, we should see every human choice as an expression of freedom; they always involve "tasting the state of freedom." In this sense, the earlier accomplishments of reason in "Conjectural Beginnings" are always

imperfect accomplishments, since they are exercises of a rational capacity whose highest form is expressed in action determined by consciousness of the moral law. But this isn't to say that the kind of problem faced by the human being who deviates from the rule of instinct could not be faced (and resolved) by a rational being whose rational capacity faced limitations that ours did not; there is no reason to suppose that the same problem could not be encountered (and resolved) by a non-free rational being.

Such a being would be like the possibility alluded to in the *Groundwork*³² for whom nature takes charge of choosing its ends, leaving reason only the choice of means. This possibility is explicitly contrasted there with the possibility of a being bound by instinct, in which nature would take charge of *both means and ends*. Although this possibility is ultimately discarded, Kant doesn't suggest that it's incoherent. Rather, the possibility is discarded in light of the fact that such a "mixed government" would be a poor arrangement, inferior to the rule of instinct, and thus in conflict with the view that nature always hits on the best possible arrangement. In a famous footnote in the *Religion*, the possibility of a non-free rational being is left notoriously open, and it seems that this being would actualize the possibility of mixed government raised in the *Groundwork*.

Given that human reason has a constitutive employment in the practical sphere, any activity that is guided by a principle that is incompatible with the moral law is the expression of an internal conflict. The previous sections of the paper tried to establish exactly this point: *given our awareness of the moral law*, any end we pursue must be represented to us as good *simpliciter*. Thus, for *us*, any end we set that cannot be conceived as good must express a way of thinking that is ultimately unsatisfactory. Moreover our capacities aren't merely on a par with those of a being whose reason is incapable of such constitutive employment in the practical realm, and whose conduct couldn't be guided by the moral law; this being would be shut out from a rational ideal. However, it doesn't follow from these points that for such a rational being there is no coherent answer to a practical question once the rule of instinct is abandoned. It's true that to the question "why should you pursue your given ends?" this being would be able to offer no answer (not even: "one has to fend for oneself" or "if I will not take care of myself who would?"). But given that his reason wouldn't be bound by the ideal that our ends withstand rational criticism, no answer would be needed (other than "this is what I find myself pursuing").

But why does Korsgaard think that this isn't a coherent possibility? Why couldn't there be a being who did not conceive of its happiness as making a legitimate claim on its will, but simply as a given end? Korsgaard thinks that the instrumentalist faces a kind of dilemma that is exemplified in cases that seem to be typical violations of PIR. Suppose, for instance, Max, who wants to go to Paris, puts quite a bit of work in taking the necessary means to get there (calls to make reservations, pay for the tickets, etc.). Max then goes to the airport, but "overcome" by fear of flights, he never enters the plane and stays home for the holidays. According to Korsgaard, the instrumentalist can understand Max's behavior in one of two ways, but both will run afoul of the claim that this is a purely instrumentalist conception of practical reason. The first way reads off of Max's hesitation a change of mind: if he did not go then the fear of flight was the more powerful force in his psyche, and thus was what he wanted most. We conclude in light of his behaviour that Max wanted to avoid a flight more than he wanted to go to Paris (at least at the moment of choice). But understood this way, PIR is inviolable: no matter what Max ends up doing, we will consider that he took the necessary means to what he most wanted. In this view:

The person's end as what he *wants most*, and the criterion of what the person wants most appears to be what he actually *does*. . . . If we don't make a distinction between what a person's end is and what he actually pursues, it will be impossible to find a case in which he violates the instrumental principle.³³

But an inviolable principle isn't a normative principle, and thus not a principle of *practical reason*. Under this view, PIR isn't the only principle of practical reason; in fact, it's not a principle of practical reason at all. We could say that under this view PIR works instead as a principle of end attribution: it allows us to read off someone's actions her ends. This move turns out to be problematic, but for now one should move on to the second interpretation of Max's behavior.

Suppose instead we say that, despite issuing in action, the end determined by his fear of flight isn't the end Max should pursue. Korsgaard suggests two ways in which this view could be cashed out: we could say either that avoiding planes isn't a rational end or a rational desire, or that avoiding planes isn't his "real" end or something that he "really" wanted. The first option is obviously incompatible with instrumentalism, since it amounts to accepting that ends themselves can be the subject of

rational criticism. But Korsgaard thinks that the same problem afflicts the second option:

If we are going to appeal to “real” desires as a basis for making claims about whether people are acting rationally or not, we will have to argue that a person *ought* to pursue what he *really* wants rather than what he is fact *going* to pursue. That is, we will have to accord these “real” desires some normative force. It must be something like a requirement of reason that we should do what you “really” want even when you are tempted not to. And then, again, we will have gone beyond instrumental rationality after all.³⁴

We can say that the instrumentalist has given us no reason why we should do what we really want, when fear, depression, etc. stand in our way. But of course, there is nothing here that should worry the instrumentalist. After all the instrumentalist should not say that we have reason to pursue anything unconditionally; we only have reason to take the means to our ends, the ends we in fact pursue, or care about. It’s a rather half-hearted instrumentalist who thinks that one has any *reason* to pursue what one wants, or any reason to adopt the particular ends one happens to have—one just *has* them.³⁵ But now it might seem that we can push the instrumentalist into a corner: for, after all, as he is leaving the airport, Max doesn’t have going to Paris as his end in any straightforward sense; going to Paris isn’t something he actually pursues, and thus this presses the instrumentalist back to the first interpretation of Max’s behavior.

Does the instrumentalist really need to retreat to this position? Why can’t the instrumentalist say that Max pursued, for instance, what *appeared* to him to be what he most wanted, but not what he really wanted? To understand this point better, it might be worth starting with a different way in which Max could foul up. Suppose as he is running to the airport at the last minute, he trips on a stray object, and can no longer walk fast enough to catch his plane. There should be no temptation here to attribute to Max the end of missing the plane or anything like that given that tripping isn’t something that Max has done intentionally. Note, however, that this possibility is enough to show that PIR could guide us, and thus be a normative principle, even if it were impossible, in any sense, to flout PIR *deliberately*. For, no doubt, Max has taken insufficient means to his ends; the necessary means would involve taking the slight detour necessary to avoid the stray object. This is enough to get us the gap we wanted to count the principle as normative: what

Max thinks is the best means to the end is *in fact* an inadequate means to this end. Max does what he *thinks* PIR requires, but he is mistaken about what it actually recommends. PIR doesn't prescribe that one should take what one *considers to be* necessary means to the ends (otherwise trying to acquire a false belief about the relation between means and ends would be one way to conform to what the principle requires), but to take what *is* necessary means to one's ends.

However, by avoiding the first horn of the dilemma this way, the instrumentalist victory seems to be a somewhat limited one. For it leaves no room for the possibility that, in the original example, Max acts irrationally when he doesn't take the plane; that he fails to do what he has reason to do. For in this case, he intentionally refrained from flying, and thus, it seems unavoidable to say that avoiding flights was what Max pursued, and so this was his end. However, it's unclear that the instrumentalist needs to accept even this limitation. For, of course, accepting that this is Max's end doesn't preclude ascribing to Max the end of being in Paris. And just as the alien, unpredictable world could obstruct his arriving in Paris, his own mindset could also prevent him from succeeding. But in the latter case, since the failure was expected, conscious, and willed, we can say that his failure is a case of irrationality. This is, indeed, what his irrationality consists in: the irrationality of adopting incompatible ends, and thus necessarily failing in the pursuit of one of them. Note that in order to make the claim that as he intentionally refrains from going into the plane, he still holds on to the end of going to Paris, the instrumentalist need not say that Max has a *reason* to pursue one end or the other; all the instrumentalist needs to do is to *attribute* these ends to Max. One can insist that we have no reason to attribute an end to an agent other than what she has a reason to pursue. But this just presupposes that instrumentalism is false; that is, it presupposes that we must have reasons for our ends, and thus it cannot be the premise of an argument against instrumentalism.

Now one might want to probe further the question of how the instrumentalist attributes ends to the agents that are neither revealed in their actions nor required by reason. There is a legitimate worry about *contemporary* instrumentalists in this neighborhood; one might suspect that out of a world of mere causal impacts, nothing will turn out that will fit the bill.³⁶ But this isn't a problem with instrumentalism *per se*, but with the empiricist notion of an end. Moreover that there is a problem

here is something that needs to be argued for; a consistent instrumentalist would think that issues surrounding the question of determining the agent's ends, or what the agent most wants, aren't issues within the theory of rationality, but rather issues that pertain to empirical psychology.³⁷

What about Kant's understanding of empiricism? Does he think the empiricist can attribute ends to agents without forcing the empiricist to presuppose anything beyond PIR? It now seems the answer to the latter question is "yes." In order to attribute appropriate ends to Max, ends that allow us to see Max as irrational, we should indeed do more than attribute specific desires to Max. We must attribute to him a reflective conception of his happiness, of an ideal of everything going in accordance to "wish and will,"³⁸ with no unsatisfied desire remaining, no sense of unease or lack, as well as a capacity for comparative judgments such that one can make assessments of relative merits of the options in terms of how well they approximate this ideal (in Kant's words, to assess their "price").³⁹ But since comparison and reflection are, according to Kant, characteristic activities of any (finitely) rational being, we can ascribe to every rational being a representation of the ideal of happiness. The ideal of happiness is, as Kant points out at various places, indeterminate. Only experience can tell us what is included in this ideal, and even experience can provide us only with a rather limited grasp of the content of our happiness.⁴⁰ Indeterminate as it is, happiness can be attributed as an end to all finite, reflective beings, and thus to all (finitely) rational beings. Moreover given that the ideal of happiness is just an ideal of a "maximum of well-being,"⁴¹ we can assume that insofar as one's given desires and inclinations are concerned, this is what a finitely rational being most wants. And this leaves open the possibility not only that one will fail to pursue what one most wants (one's happiness) because one is ignorant of what it consists in, but also that one will fail to pursue it, even when one is fully aware that this is what one is doing. Since nothing we said on this issue depends on the conception of happiness as making *legitimate* claims on our will, it's fully available to the empiricist. Thus empiricism, on Kant's conception of it, can ascribe to Max the end of going to Paris as that which he most wants (as something he is aware is a constituent of his happiness), and can thus ascribe irrationality to Max in the same way as the Kantian can.

However, this position doesn't leave instrumentalism unscathed. For part of the appeal of the instrumentalism, and of

empiricism in its most common contemporary guise, lies in the fact that it starts from a seemingly unassailable fact: our happiness obviously matters to us. The fact is indeed unassailable, but also unavailable to the empiricist who doesn't have at her disposal any ordinary conception of happiness. The empiricist version of happiness seems to be one of the general objects of blind attraction rather than the general object of an *intelligible pursuit*. A life spent howling at the moon in this conception is in principle just as qualified to be what one's happiness consists in as anything else. This is, for Kant, not the notion of happiness that figures in ordinary thinking.⁴²

This reading also suggests, somewhat indirectly, another related difficulty that instrumentalism faces. The fact that happiness could be a given end for a rational being, an end given by its empirical nature, rather than set by its own reason, seems unproblematic to Kant.

For Kant, the fact that some ways of living are happier than others, and the fact that experience sharpens our understanding of what happiness consists in is quite obvious. But when we examine how Kant actually tries to determine the object of happiness, in particular in the *Anthropology*, we cannot avoid the feeling that this is a quaint conception of the nature of empirical inquiry. We would not expect to find the following passage in any scientific treatise on any aspect of human nature (let alone one that would claim to belong to a discipline called "anthropology"):

The good living that seems to harmonize best with virtue is a good meal in good company. . . . At a full dinner, where the multitude of courses is only intended to keep the guests together for a long time, the conversation usually goes through the three stages of 1) narration, 2) reasoning, and 3) jesting. . . . In the third stage . . . [t]he conversation turns naturally to mere play of wit, partly also to please the lady in the company who is encouraged by the minor, intentional, but not insulting attacks on her sex to shine in her own display of wit.⁴³

The problem of this passage isn't the content of the advice: it's not that we may find dinner parties frivolous, or that we may doubt that these kinds of attacks would always be delightful for the lady in the company. The problem is that we cannot make clear sense of an empirical science that purports to give this kind of advice on good living, whereas its suitability for this kind of advice is what makes anthropology for Kant a worthy scientific enterprise.⁴⁴ Anyone making such claims would find her work more

likely to be shelved in the self-help section than among respectable scientific treatises. If modern empiricists take empirical inquiry to be modeled by the sciences,⁴⁵ Kant's untroubled acceptance that experience can determine what happiness consists in cannot be endorsed without further work. One must show that a suitable notion of happiness or satisfaction, or in general, a suitable notion of a given end is indeed available from empirical sciences such as, for instance, empirical psychology.

Given our understanding of the nature of empirical science, can modern empiricists have a coherent conception of a given end? Bereft of our ordinary conception of happiness, can we find a mental state that would enable us to perform the same task? And if so, why should one transfer one's attitude to one's happiness ordinarily conceived (the thought that our happiness obviously matters to us) to the objects of this state? Our contemporaries, unlike Kant's contemporaries (as he read them) claim not only a practical advantage but also a metaphysical advantage for an instrumentalist conception of rationality. Not only does our happiness obviously matter to us, but we can also appeal to subjective ends without incurring any metaphysical commitments beyond those of a scientific empirical psychology. But the above questions suggest that metaphysical and the practical advantage might be in tension. Indeed one might start suspecting that the following two features cannot be coherently combined:

(1) The criterion of happiness (or satisfaction) is provided by mental states that are discoverable by a scientific psychology

(2) Other things being equal, it's better for someone to be happy (or to be satisfied, etc.) than to be unhappy (or to be dissatisfied, etc.).

Since Kant did not share with us (or at least with contemporary empiricists) a conception of what empirical psychology must look like, there is no reason for him to suspect that empiricism might be an incoherent position solely on account of having to come up with a conception of an end that is merely given. Still, the suspicion remains that there is no coherent conception of happiness answering to these two claims. In the same vein, contemporary instrumentalists might try to argue that happiness *as they conceive* it is all that could matter to us. But the move from the fact that obviously our happiness matters to us to optimism about the availability of such a notion to contemporary instrumentalism is questionable.⁴⁶

NOTES

References to Kant's works are to the standard *Akademie* pagination, with the exception of *Lectures on Ethics*. References to the Lectures on Ethics are to *Eine Vorlesung über Ethik*, edited by Gerd Gerhardt (Frankfurt am Main: Fischer Verlag, 1990). Specific works are cited using the abbreviations below. I have used the English translations below with occasional minor changes.

- APH *Anthropology from a Pragmatic Point of View*, trans. Victor Lyle Dowdell (Southern Illinois University Press, 1978).
- G *Grounding of the Metaphysics of Morals*, trans. Mary Gregor (Cambridge: Cambridge University Press, 1997).
- KpV *Critique of Practical Reason*, trans. Mary Gregor (Cambridge: Cambridge University Press, 1997).
- MAM *Conjectural Beginning of Human History* trans. Ted Humphrey (Indianapolis: Hackett Publishing Company, 1983) under the title *Speculative Beginning of Human History*.
- MS *Metaphysics of Morals*, trans. Mary Gregor (Cambridge: Cambridge University Press, 1991).
- R *Religion Within the Boundaries of Mere Reason*, trans. Allen Wood (Cambridge: Cambridge University Press, 1999).
- VE *Lectures on Ethics*, trans. Louis Infield (Indianapolis: Hackett Publishing Co., 1981).

1. Christine Korsgaard's work has been particularly influential in presenting Kantian arguments against the coherence of instrumentalism. See "The Normativity of Instrumental Reason" ["Normativity"] in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (Oxford: Oxford University Press, 1997); and *The Sources Of Normativity [Sources]* (Cambridge: Cambridge University Press, 1996).

2. It is important to note that not all philosophers who fall within this category would be naturally described as "empiricists." Wolff's practical philosophy, for instance, would, for Kant, fit this description. See KpV, Remark 2 to Theorem IV.

3. At KpV:22, Kant identifies the "principle of self-love" with "the principle of one's own happiness." Kant also refers sometimes to the "principle of happiness" (See, for instance, KpV:36). Since for our purposes, any differences that there might be between these two principles are irrelevant, I will refer to either of these principles simply as PH.

4. KpV:35, emphasis added.

5. G:405.

6. G:405.

7. KpV:35.

8. KpV:35–36.

9. KpV:36.

10. KpV:23.

11. See Andrew Reath, “Hedonism, Heteronomy, and Kant’s Principle of Happiness,” *Pacific Philosophical Quarterly*, vol. 70 (1989), pp. 42–72.

12. *Characteristics of Men, Manners, Opinions, Times*, ed. Lawrence E. Klein (Cambridge: Cambridge University Press, 1999), p. 202, italics added.

13. KpV:23.

14. R:36, emphasis added.

15. VE,222. I discuss this issue in more detail in “Friendship and the Law of Reason,” in *Persons, Promises and Practices*, edited by Joyce Jenkins and Christopher Williams (Notre Dame, Ind.: University of Notre Dame Press), forthcoming.

16. Cf. R:36.

17. See, on this issue, Stephen Engstrom, “The Concept of the Highest Good in Kant’s Moral Theory,” *Philosophy and Phenomenological Research*, vol. 52 (1992), pp. 747–780.

18. KpV:34.

19. See KpV:110.

20. See KpV:106 for this definition of “good.”

21. KpV:74.

22. R:42.

23. See KpV:61.

24. G:405.

25. See, for instance, Korsgaard, “Normativity,” and *Sources*.

26. KpV:24.

27. See, for instance, Philippa Foot, “Are Moral Considerations Overriding?” in *Virtues and Vices* (University of California Press, 1978).

28. See *Sources*, chs. 1 and 2, and “Normativity.”

29. For a reading of this passage roughly along these lines, see Tamar Schapiro, “What is a Child” in *Ethics* 109, 1999, pp. 715–738.

30. MAM:112.

31. Cf. MAM:113.

32. G,395.

33. “Normativity,” p. 230.

34. *Ibid.*

35. Korsgaard seems to think that I can infer from (1) “X is my end,” and (2) “Y is the necessary means for X,” to the conclusion that (3) “I ought to pursue Y.” This is a mistake. The instrumental principle (at least as understood by instrumentalists) is a principle of consistency, and thus the consequent isn’t detachable. It should be read not as “If one pursues the end, one ought to pursue the necessary means,” but as something like: “It ought to be the case that if one pursues the end, one also pursues the necessary means.” See on this issue, John Broome, “Normative Requirements,” in *Ratio*, vol. 12 (1999), pp. 398–419; and “Practical Reasoning,” unpublished manuscript.

36. More on this issue below.

37. Or perhaps a matter for some other kind of study of human nature. See on this issue, Philippa Foot, *Natural Goodness* (Oxford: Oxford University Press, 2001).

38. KpV:124.

39. Cf. G:434–435.

40. See, for instance, G:418.

41. G:418.

42. At least some contemporary philosophers accept some version of the claim that even desire must have an intelligible object. See, for instance, G. E. Anscombe, *Intention* (Oxford: Basil Blackwell, 1957), Warren Quinn, “Putting Rationality in Its Place,” in *Morality and Action* (Cambridge: Cambridge University Press, 1993).

43. AHP 278, 280–281.

44. Cf. AHP 119–120 when Kant compares the aims of the *Anthropology* with those whose inquiries place them as “mere spectator[s].” The latter kind of inquiry Kant considers to be a “sheer waste of time.”

45. No doubt, modern empiricists might deny that empirical enquiry should be identified with scientific enquiry. See, for instance, David Copp, “Why Naturalism,” *Ethical Theory and Moral Practice*, forthcoming. But in this case, it’s unclear whether their view has any metaphysical advantages over rationalist views.

46. For extremely helpful comments, I would like to thank an HPQ referee, Donald Ainslie, Arthur Ripstein, audiences at Carleton University and the 23rd NEASECS, and, especially, Jennifer Nagel.

