

The Self-Sufficiency of Instrumental Reason. A Defense of Kant's Theory of Hypothetical Imperatives

I.

An agent has an instrumental reason to perform an action A or to try to achieve an end E, if A and E are a means by which she can achieve X, which she wants to achieve. Instrumental rationality (or instrumental reason) consists in *applying* the appropriate means to achieve what you want to achieve. We have to distinguish between two questions here:

First, (there is) the question of *how* something can be achieved that one wants to achieve. In other words, this is the question of *what* the appropriate means are to achieve something. This is a theoretical question, as it is meant to determine what is the case. Can an end be achieved by doing this or by doing that?

Second, we have to deal with the question of whether someone has a sufficient reason to actually apply the appropriate means. This is a practical question, since it is meant to determine how someone ought to act.

By pointing to the thesis of self-sufficiency of instrumental reason, I suppose that only from the fact that an agent actually wants to achieve something, it may follow that she ought to apply the means necessary to achieve what she wants to achieve. There are genuine instrumental norms of action, and there would also be instrumental norms of action, if there were no other kinds of norms of action. By arguing in favor the thesis of self-sufficiency of instrumental reason I do not want to assert that there are only instrumental norms of action. I do not want to deny that genuine(ly) moral norms of action, if they exist, have priority over instrumental norms of action and that they can restrict instrumental norms of action accordingly. Still, the existence of genuine moral norms of action needs to be shown separately. There can be instrumental norms of action without the existence of moral norms of action.

My thesis of the self-sufficiency of instrumental reason is completely in accordance with Kant's theory of hypothetical imperatives. In Kant's view, hypothetical imperatives are conditional norms of action. They are contingent on the presumed or possible ends of action. They indicate that an agent ought to apply the means necessary to achieve what she wants to achieve. They are in force for the respective agent as or insofar as she has certain ends. These norms of actions are 'possible', as Kant says, i.e. they possess a necessitating force for an agent. There can be no doubt about the existence or reality of

these norms of action. The question is rather whether there are also unconditional norms of action, i.e. categorical imperatives. It is questionable whether such unconditional norms of action may possess a necessitating force on us.

But Kant does not only assert the reality of conditional or instrumental norms of action, he has also shown how they come about. In what follows I would like to reconstruct his argument and show that it is successful. In my opinion, from a systematic point of view, this is an extremely significant and yet unfortunately unappreciated piece of the theory of Kant's practical philosophy which is very often misunderstood. Therefore, I would like try to explain and defend Kant's theory of hypothetical imperatives.

II.

For Kant, technical imperatives, or technical norms, as we would call them today, are the model case of hypothetical imperatives. They indicate that if one wants to achieve an end E, one has to use the means necessary to achieve this end. In order to see how the *fact* that an agent wants to achieve something can establish a *norm* that prescribes what an agent ought to do, it is necessary to take the perspective of the agent who wants to achieve a particular end.

I would like to explain the coming-to-be of a technical norm by giving you an example. Suppose I want to lose 20lb. This is not just a desire that I have, but I am really determined to lose weight. Losing weight represents a real end for me. But it can only be a real end, if I also think of how I can achieve it. Suppose I come to the conclusion that the only way to achieve my end consists in completely giving up the pleasure of chocolate without substitution. Now, it may be possible, that this is not an option for me. I just cannot stop eating chocolate. If this is the case and if my end of losing weight cannot be reached by other means, then I will have to abandon my end of losing weight. The reason for this is that I cannot be consistently determined to reach an end, and not be willing to use the means necessary to achieve this end. But if I do not immediately reject to apply a means necessary, my wanting to reach the end of losing weight makes it necessary to want to apply the means to stop eating chocolate without substitution.

Thus, really wanting to achieve an end is accompanied by wanting to apply the necessary means to achieve it. Therefore, Kant speaks of an "analytic practical sentence". For Kant a "sentence" signifies a "judgment". Accordingly, a sentence is a practical judgment.

Thus, by wanting to reach the goal of losing weight the practical necessity arises for me to also want to apply the means necessary. Now I do not automatically want to do what is necessary for me to want to do on the basis of my actual end. Time and again I rather feel a strong urge to eat chocolate, i.e. to do the contrary of what would be practically necessary for me to do. Therefore, the practical necessity of wanting to do something, arising out of my actual ends, confronts me as a demand, a necessitation, an ought. Since it is a particular kind of demand, namely the demand to use the means necessary to achieve my end, it might be useful to characterize it as a technical “ought” and to index it accordingly: "I technically ought_T to eat no chocolate (now)."

I may react to this demand in different ways in the form of specific practical judgments. Maybe I will follow the norm. Maybe I will tell myself that I will allow myself to make an exception. I will have some chocolate now, but I won't have any more chocolate in the future. Maybe I will abandon the end of losing weight as it leads to such a bad norm not to eat any chocolate any more. The abandonment of the end comes with the disappearance of the demand.

Kant's theory of technical norms is important because it provides a theory of the emergence or formation of practical norms. Practical norms are generated by our will. For technical norms this can be shown in the form of a sequence of necessary judgments of an agent. You can find it on the handout.

- (1) „I want (to achieve) end E.
- (2) „I am positively determined to accomplish or achieve E.“
- (3) „I assume that there exists (at least) one means M the application of which is necessary for the accomplishment or achievement of E.“
- (4) „I assume that it is within my power to apply M.“
- (P₁) „Through this M = X, the application of which is in my power, E can be achieved.“
- (5) „I must (the necessity exists from my point of view to) apply this M = X.“
- (6) „For me, the necessity exists of wanting to apply this M = X.“
- (P₂) „I ascertain that it is not self-evident that I want to apply this M = X.“
- (7) „I (technically) ought_T to apply this M = X.“

Judgment (2) does only interpret judgment (1), i.e. it makes explicit what judgment (1) as a judgment of an agent amounts to. For a correct understanding of the sequence it is important to bear in mind that judgment (1), as well as the other judgments, is temporally indexed. The sequence holds if and as long as the agent upholds his resolution to achieve end E. The wanting of E can, of course, not bring forth by itself the availability of suitable means. But the agent cannot consistently hold that there is no means available whatsoever, either in general or in the concrete situation, by which he might achieve the end and at the same time uphold judgment (1). At least in this sense, the judgments (3) and (4) are necessary judgments for an agent who makes judgment (1). Thus, the agent must make a judgment which determines a means that is appropriate to achieve the end and whose application is in his power. This judgment is a theoretical judgment and the agent may err in the determination and assessment of the means. In this regard, the concrete judgment is not part of the sequence of necessary judgments. Instead, it functions as a sort of premise for the further judgments. Therefore, I do not count it as a further judgment of the sequence but label it as P_1 . On the other hand, the agent is forced to make a judgment of the kind of P_1 as a consequence of judgments (3) and (4). While P_1 is a theoretical judgment, judgment (5) is a practical judgment insofar as it articulates what the agent has to do. Judgment (6) interprets or explicates judgment (5). The next judgment, which I label P_2 , is again not a direct part of the sequence but is important for its continuation or completion. It articulates an experience of the agent in the face of the practical necessity of wanting to employ the necessary means. The agent experiences a resistance against wanting to do what he is forced to do. Hence, the practical necessity to want to employ the means confronts him as a demand or an ought. Thus, provided that the agent has to accept judgment (5) or (6) together with judgment P_2 , it follows that judgment (7) is a necessary judgment for the agent. The ought in judgment (7) is a technical ought which is dependent and conditional on the agent's wanting to achieve end E. The agent can always get rid of the demand to apply $M = X$ either by applying M in such a way as is necessary to achieve the end E or by giving up end E. In order to achieve an end it may be necessary to carry out various actions and, if necessary, to do so in different and coordinated ways. In order to achieve the end to lose weight, one has to eat less for weeks (which is one of the reasons why it is so difficult to lose weight.)

Real ends lead to actual technical norms. But there can also be ‘virtual’ technical norms that will apply only *if* an agent has certain ends. Manuals or instructions for use and recipes are examples for such virtual technical norms.

III.

As sentient beings we want to achieve and maintain pleasant conditions and we want to end and avoid unpleasant conditions. By means of our reason we form the idea and the end of a state of convenience in as comprehensive a way as possible with as low a degree of inconvenience as possible or the idea of the greatest possible well-being or happiness for oneself. Kant refers to the goal of one’s own happiness as a "natural purpose". This expression is deliberately paradoxical. For purposes have something to do with choice and freedom. In contrast to that, nature implies something like predetermination and inevitability. The objective of our own happiness is an objective which is built up as if it were made behind our backs and which is inevitable as we are sentient beings capable of reason. We have never made the decision to want to be happy. We always (already) want to be happy. Accordingly, the end of one’s own happiness does not represent a specific objective. Rather, it is, as Kant points out, a vague and diffuse objective.

This vague and diffuse objective can guide the choice of our ends. We have to continuously define it by the development of overarching objectives, without ever being able to definitely or exhaustively interpret it. Therefore, each of us has to interpret it individually and there can be no guarantee that our interpretations will help us to achieve the objective of our own happiness. There are experiences of what usually amounts to happiness and what is usually detrimental to it. But there are no guarantees. Health, a fulfilling partnership, a certain degree of prosperity and a rewarding profession will normally contribute to the objective of our happiness. If we have reason to assume that the relevant overarching ends will contribute to our happiness we have reason to pursue the relevant ends and try to achieve the appropriate sub-goals. From this perspective, my aim of losing weight may represent a sub-goal, which is necessary for me to achieve in order to reach my overarching end of long-term health. I ought then not just give it up.

Even if the relations between means and ends are not stringent (Can nonsmokers really be happy? Doesn’t an early heart attack save a person from long hardship and infirmity?), we are nevertheless dealing with practical demands here which in principle have the same structure as technical norms. Since I want to achieve something, the prac-

tical need arises for me to want to achieve something else. As I do not automatically do what is necessary for me to do, I'm facing this practical necessity generated by my will as a demand, it becomes something I ought to do. Since the respective demands are about my considered self-interest, which means that they are demands of prudence, we should also index them, such as in a judgment like "I prudentially ought_p to try to lose weight".

Although prudential norms arise from the same instrumental rationality as technical norms, they have precedence over technical norms in the case of a conflict. This is due to the overarching purpose of one's own well-being, particular ends can be judged and criticized with. But prudential norms of prudence do not simply void technical norms. It remains that, if one wants to achieve an end E, one technically ought_t to apply a particular means or one of the available means. But one prudentially ought_p either not to try to reach end E or not to apply the available means. In such a case, the technical norm as a technical norm is not wrong, but it is not pertinent. The same would apply for norms of prudence, if there is an end or if there are ends for agents which are able to limit the natural purpose of one's own happiness. Even then it would still be prudent to pursue certain objectives, but the agent morally ought_M not always do what she prudentially ought_p to do..

The possible limitation of the purpose of one's own happiness by a higher or supreme end is not to be confused with the limitation of self-interest out of self-interest. Already in the second book of Plato's *Republic*, the norms of coexistence in a society are explained by the self-interest of those affected. While we would like pursue our own self-interest without limitation, we also do not want others to fully pursue their own self-interest. Therefore, we are all better off if we are mutually willing to restrict our own self-interest and to accept mutually restrictive norms of coexistence. As regards content, these norms look like moral norms. But, based on this rationale, they remain prudential norms. Therefore, it may always be the case that individuals have no sufficient reason to follow these norms, and this is the case, when it is in his or her considered self-interest to violate one of these norms.

It is highly controversial to what extent moral norms can be reduced to norms of prudence, as regards content. But even if the corresponding attempts carried out e.g. by Thomas Hobbes, James M. Buchanan and David Gauthier should fail, this would not imply that certain norms of action cannot completely be justified instrumentally. Kant's theory of hypothetical imperatives which I have outlined briefly, shows that a purely

instrumental justification of norms of action is possible and that such instrumentally justified norms of action are real.

IV.

I have explained that Kant's theory of hypothetical imperatives is a theory of the emergence or formation of norms. Norms are created by us through our own will. This can also be expressed as follows: Again and again, an agent bridges the gap between "is" and "ought" in and through her will. Her practical judgments compel an agent to recognize practical necessities and demands over and over again. These necessities and norms solely result from requirements of consistency. Reason does not provide genuine orientations in the case of hypothetical imperatives. In the case of hypothetical imperatives, reason does nothing more than "to look after the interests of the inclinations" (GMS 406), according to Kant.

For a contemporary philosopher, especially from the English-speaking world, it seems to be almost impossible to comprehend which judgments a (human) subject of cognitions and practical decisions can make and has to make. (Thus it seems to be almost impossible to take on the perspective from which the works of early modern philosophy like those of Locke, Berkeley, Hume and Kant are mostly written). Instead, the semantic implications of syntactically well-formed sentences is asked for. The sentence "I want to achieve end E" semantically does not contain an "ought". This is also the reason why Thomas Hill finds it necessary to presume that Kant applies a normative principle he refers to as "The Hypothetical Imperative".

According to Hill, this principle can be tentatively formulated in the following way: "If a person wills an end and certain means are necessary to achieve that end and are within his power, then he ought to will those means."¹ It should have become clear from my sketch of his theory of hypothetical imperatives that Kant does not require such a principle. Kant would also have totally rejected such a principle. For there immediately arises the question of the status of the "ought" of this principle, and accordingly, of how this principle is justified. According to Hill it complies with the "ought" of the categorical imperative. By doing this, Hill does not only make the norms of instrumental reason disappear, but he also has to represent an entirely implausible (and Non-Kantian) theory, which tries to solve the self-created problem of conflicts between "The Hypothetical

¹ Thomas E. Hill, Jr., *The Hypothetical Imperative*, in: *Philosophical Review* 82 (1973), 429-450, 429.

Imperative" and "The Categorical Imperative". I will refrain from going into detail here. Instead, I would like to address Christine Korsgaard's criticism of the normativity of instrumental reason.

V.

In her article "The Normativity of Instrumental Reason",² Christine Korsgaard holds "that hypothetical imperatives cannot exist without categorical ones, or anyway without principles which direct us to the pursuit of certain ends, or anyway without *something* which gives normative status to our ends." (p. 250) She is of the opinion that this is also the view Kant finally adopted, but had not yet reached in the *Groundwork of the Metaphysics of Morals* from 1785, which contains Kant's theory of hypothetical imperatives. This theory of the commanding character of hypothetical imperatives according to Korsgaard still shows traces of a dogmatic rationalism and must therefore fail.

Korsgaard discusses hypothetical imperatives under the designation "instrumental principle". In the context of her discussion of David Hume's position, she offers two alternative formulations of the principle:

"(a) 'if you have a *reason* to pursue an end then you have a reason to take the means to that end.'

(b) 'if you are *going* to pursue an end, then you have a reason to take the means to that end.' " (p. 223)

According to Korsgaard, the formulation (b) of the principle is problematic, because it derives a normative conclusion from a fact. This is not the case with formulation (a). Here, the normative, i.e. the reason to employ the means, depends on something which is superiorly normative, namely the reason to achieve an end. Because the superior reason cannot be instrumental or at least ultimately not be dependent on an instrumental reason, the instrumental reason articulated in the principle cannot be independent. "[I]f you hold that the instrumental principle is the only principle of practical rationality, you cannot also hold that desiring something is a *reason* for pursuing it." (p. 223)

But the formulation (a) of the instrumental principle is ambivalent. For the reason one has to pursue an end can be either an "objective" or a "subjective" reason. I may have a subjective reason for pursuing an end without having an objective reason or

² Christine M. Korsgaard, *The Normativity of Instrumental Reason*, in: Garrett Cullity/ Berys Gaut (eds.), *Ethics and Practical Reason*, Oxford: Clarendon Press, 1997, 215-254.

without conforming to an objective reason which perhaps exists. For my reason to pursue an end can consist only in wanting to achieve something more than something other. To put it more generally: An animal may be hungry and automatically or instinctively start the search for food. I feel hunger and perhaps *decide* to eat something. My decision is a resolution to pursue an end. Normally, I will bear a concrete action ϕ in mind which I have to carry out in order to achieve my end E. At least implicitly my decision to pursue an end has a preferential structure. I hold it to be better to pursue E than not to pursue it. The criteria on which my preference is based may be of various kinds. I may prefer E because I expect to derive pleasure from it or because it contributes to my considered self-interest or because I think that it is morally obligatory to pursue E. It is important to note that my concrete decision to pursue or not to pursue E is determined by what I *hold* to be better and not necessarily by what is ‘objectively’ better and what I *should* therefore hold to be better. I may even know that my action will probably be contrary to my considered self-interest or will be morally wrong and that therefore I ought not to do it. In such a case, such considerations are not relevant or decisive for my decision. Possibly, I will revise my decision or even regret it at a later point in time. But this does not change the fact that the reason for my decision to pursue an end might not be based on what (according to what criteria soever) I ought to do but on what I actually want to do (even if I ought to do something different). Such a reason is what I call a “subjective” reason. This has to be distinguished from an “objective” reason which refers to standards or norms (which may derive from the agent himself or from other sources). Now, I pointed out that technical norms or hypothetical imperatives derive from the (actual or, in the case of virtual norms, tentatively supposed) decisions to pursue an end. Thus, if the decision to pursue an end can, in the sense just explained, be based on a “subjective” reason, then Korsgaard’s claim that the normativity of instrumental reason cannot be independent or autonomous is wrong.

Not unlike Thomas Hill, Korsgaard interprets Kant’s argument as a deductive inference. In doing so, she takes “Whoever wills the end wills the means insofar as he is rational” as the first premise of the following syllogism:

“Whoever wills the end wills the means in so far as he is rational. I will the end. → Therefore I will the means in so far as I am rational. → Therefore I *ought* to will the means.” (p. 239)

One need not be astonished that Korsgaard holds this syllogism to be unsatisfactory:

“But we cannot in any non-trivial way invoke this [...] syllogism to explain *why* the agent finds it rationally necessary to take the means to this end, for this syllogism’s first premiss trivially incorporates the claim that taking the means to one’s ends is rationally required.” (ibid.)

But, as I showed, Kant does not presuppose a norm of rationality which is independent from the agent. Instead, he shows that conditional norms or hypothetical imperatives are demands which are based on a practical necessity, which is a consequence of a requirement of consistency for the agent. The agent cannot consistently want to pursue an end and at the same time not want to employ the means necessary to achieve the end.