KANT AND ARTIFICIAL INTELLIGENCE: A REVIEW ESSAY

Kant and Artificial Intelligence
Eds. Hyeongjoo Kim and Dieter Schönecker
Walter de Gruyter, 2022

Randall E. Auxier
Southern Illinois University Carbondale (USA)
personalist61@gmail.com

Pragmatists and Kant

There is a problem that pervades Kant interpretation which needs to be set right. It has to do with the way pragmatists (and others) interpret his view of "necessity," which is far more subtle and important than the oft-rehearsed formulas drawn from the *Groundwork of the Metaphysics of Morals* and *The Metaphysics of Morals* –works concerned exclusively with the *form* of the will, not with the will as enacted *empirically*. One assumes (wrongly) that the form of the will subsumes the empirical will, and using the logic of Reason (alone), it does. But there is far more to the logic of action than can be exhausted by the logic of pure practical reason, i.e., those things we can think about relative to action, and the norms governing how we *ought* to think about such matters. Thinking is only one kind of acting, however, and it lacks the power fully to determine our action. We can only say that when we have adhered to the rules that ought to govern our thinking about action, we have a clearer idea of what we should do.

As a pragmatist I am not sympathetic to any claims of necessity in which some sort of logic motivates some sort of claim about how the world (moral, natural or supernatural) *must* be. But neither is Kant sympathetic to such claims, as too many pragmatists forget, or never learned. It is important to remember that Kant gives us the word "pragmatism," and that his anthropological writings (flawed as they are), as he explicitly says, are empirical and therefore fallible, subject to change over time as we learn more about the human story. The main thing to remember, as that anthropological story unfolds, is that *ought* still implies *can*. This applies to the norms of thinking as much as it does to the norms of action (thinking is, after all, acting). If I *ought* to think in this way rather than that about the relationship *between* my thinking and my other actions, then I can rest assured that I *can* think about it in one way rather than the alternatives. If I can't think about, for example, future empirical discoveries that haven't been made, then there is no force in saying I ought to, or ought to be responsible for forming my will around the universalizability of the maxim taken from such actions.

That brings us to "artificial intelligence," so called. We are justified, I argue, in updating the empirical side of Kant's anthropology to address the questions raised by AI. Note that I do not accept this label. It is not "artificial," if that is supposed to denote "other than natural," especially in Kant's sense of nature. Rather, it is a perfect instance of what Kant calls "mechanical" and as such best understood as under the domain of nature. Nor is AI "intelligent" in Kant's sense of that term. I cannot here go into the details of these matters of how to label, but I can recommend some other writings that illuminate the subject.[1] But with this said, it is clear that we, today, need to think about the implications of LLMs and other mechanisms that go under the label "AI" in ways that do take account of our duties to others and ourselves. I am not a deontologist, but then, in my view, neither was Kant. The story is not as simple as the one told by the last hundred years of duty-ethics. It is not as if pragmatists would advise us to *ignore* the idea of duty.

What, then, is our duty with regard to inquiry, thinking and acting, in the presence if LLMs and these other mechanisms? We must be pragmatic. There we hit something that Kant would have known, since he invented the idea. It seems that if Kant makes something obligatory (for thinking or other actions) but we lack the neces-

[1] See Auxier and Mueller, "Kant, Moral Imagination, and the Pathologies of Reason," with Laura J. Mueller, in Studia Philosophica Wratislaviensia, Vol. XVII, fsc. 4 (2022), pp. 5-27. https://wuwr.pl/spwr/article/view/14992.

sary means of performing the act (including thinking the thought), then his own formula of "ought implies can" swoops in to save the day. That principle is categorical for Kany, not hypothetical, as with Hume, and means that the act must be *possible* "under natural conditions" (C1 A548/B576). Please do note that qualifier –under natural conditions. Whatever does this mean?

Unlike most pragmatists, I am not allergic to Kant. Kant is not, however, a communicable disease in a room of philosophers immunized by pragmatism. They don't think they have to consider his ideas. So I am lonely sometimes, in my infirmity of thinking that pragmatists do and must think about these ideas. Dewey wrote his dissertation on Kant. So did Royce. Peirce is steeped in Kant. Only James, who had no training in the history of philosophy, was allergic, and largely ignorant of this philosophy. It is not good for pragmatists to follow that path. They persist in a misunderstanding shared with analytic philosophers, and born of laziness in reading. The last 120 years of ethics has drawn on senses of "necessity" that collapse a number of important distinctions Kant was using in his deployment of *Notwendigkeit*. It makes him a bit mystifying to read and follow these days, especially in English, where a number of subtleties are lost in translation.

I am pleased to say that the book I want to consider here *Kant and Artificial Intelligence*, edited by Kim and Schönecker is a corrective to these problems, if pragmatists would read it. The editors and contributors on producing an important volume in the long history of philosophical discussions of Kant. As several contributors point out (and I have emphasized above), the idea of artificial intelligence did not exist when Kant was alive, so speculation is unavoidable, and updating the critical system for application to the new problems is obligatory, for our thinking and other actions. However, as they also rightly say, a great deal of what Kant said bears on this topic, and there can be no doubt that Kant's philosophy has had a huge effect on how the debate on artificial intelligence has unfolded, from Turing to the present. People have used his ideas, but perhaps not in the way pragmatists would and should, due to the unfortunate allergy.

This volume is not bound by the same standards that high-end Kant scholarship must meet, since "getting Kant right" is not a categorical imperative for this kind of inquiry. In order to get to the main business of the book the contributors must draw on their own readings of Kant. Nevertheless, for the sake of clarity we should divide the effort into three kinds of attitudes: (1) there are the applications of Kant which seek to stay within some chosen "accurate" tradition of Kant interpretation, remembering that there are four or five major strains and several minor branches of Kant interpretation. (2) There are those efforts that seek to build outward from Kantian ideas but without any special loyalty to the text or scholarly traditions descending from Kant. These might be called Kantian in spirit but not according to the letter. (3) There are efforts that do not focus on Kant in any special way but rather treat Kantian philosophy as a toolbox for approaching contemporary issues in their own terms, and which therefore do not seek to be even Kantian in spirit. We find all three attitudes in this volume. And perhaps we would find most pragmatists in the third group.

I want to assess these essays on their own terms rather than quarrel with whatever version of Kant they put forward. I don't want to impose purposes on them that are foreign to their attitude. Yet, I have my own ideas about how to handle Kant on this topic, and also how to approach the philosophical issues of artificial intelligence. It is difficult to set one's own perspective aside, and so I will include mine instead of pretending to an objectivity I doubt anyone really possesses. As I worked through the volume, I slowly realized that my own views about how both Kant and AI should be handled could not be wholly left out.

Getting Kant Right

Therefore: I should make it clear that as an interpreter of Kant I am a defender of the full architectonic, which means that I regard most English language Kant interpretation as irrelevant to Kant's actual purposes and accomplishments. It also means that I part ways with most pragmatists regarding Kant because I don't think he has been rightly understood in our tradition. The naturalistic and analytic readings descending from Kemp Smith, and bent to the purposes of 20th century logic are, quite simply, not Kant, and even historical interpreters trained in the analytic tradition are usually bad interpreters, since their ideas about logic, truth, knowledge, and the like are almost entirely foreign to Kant's ideas about these topics. Making Kant relevant to or respectable to and by analytic philosophers does not lead to good interpretation. The first generation of pragmatists knew very well not to give way to the logic of Russell and his hoarde of simplifiers. But somewhere in the confusion, the pragmatists lost their understanding of better logic, better epistemology, and better metaphysics than these narrow ideas can provide. Instead of inheriting and improving on Kant, we cut off our faces despite our noses. There are exceptions to this summary, but most of the recent Kant scholars, even some very prominent ones , and nearly all pragmatists are, in my opinion, deeply misguided. Most do not accept the architectonic on its own terms, and make too much of a distinction between Kant's pre- and post-critical thought.

My own reading is closest to Cassirer's, and I regard the critical phase of Kant's thought as more continuous with the pre-critical phase than is typical for English language Kant scholarship. I am impatient with people who don't finish the first *Critique*, which is much of the English-speaking world, and who treat Kant's moral philosophy as a detachable and independent ethical theory, and with those who don't even read the third *Critique* –who don't understand that logic is about judgment, and who oversimplify To understand Kant requires the study of his own logic, and its differences from contemporaries, such as Christian Wolff and Frederic Castillon. Without this effort, one will never understand the structure of Kant's arguments. He does not employ the logic of the 20th century, and would not use the so-called "mathematical logic" even if he knew it. The idea of divorcing logic from the *act* of judging would make no sense to him, nor would the idea of collapsing all judgment into a simplistic judgment of true or false. Judging is something we do. Judgment is something we *think* about, theorize, as well as do. Yet, Kant was perfectly comfortable with mechanizing our thinking and processes of *describing* our judgment, and he advised pressing that strategy as far as possible, but in no sense can we achieve complete explanations thereby. This relative comfort with mechanization is, of course, relevant to the question of the book, and it does come up. But the limitations Kant places on this strategy carry the day.

> Kant says:
> Judgment, through its a priori principle of judging nature in terms of possible particular laws of nature, provides nature's supersensible substrate (within as well as outside us) with *determinability by the intellectual power.* But reason, through its a priori practical law, gives this same substrate *determination*. This judgment makes possible the transition from the domain of the concept of nature to that of the concept of freedom. (C3 196 Pluhar translation, emphasis in the original).

In short the things in themselves are determin*able* by our thinking, and our reason, in virtue of its practical powers, can determine for our thinking the noumenal realm made available to us by our postulates of pure practical reason. This means that we are capable of thinking about possible actions (both inside and outside us) as "can" –as possible. Pragmatists need this power as much as anyone –think of Dewey's theory of dramatic rehearsal of actions we might perform, since we cannot think in situations when action must be taken without delay. What Kant is saying is not so different.

This transition from judging to judgment is subjective, because the determination involved is the subject

determining *itself* in the mode of a feeling of liking or not liking, but this power of determination is not limited to judgments that "x is beautiful"(which is only an illustration of reflective judgment), it us a characteristic of all judgments. The reflective judgment is weak, has little influence over action, but it is comprehensive, i.e., inclusive, of the determinate judgments of reason and understanding, and allows us to see each in light of the other –particular causal laws in light of freedom, freedom in light of limiting circumstances in nature.[2] All of this is importantly relevant to questions of whether mechanized inference is or is not "intelligent."

What is necessary, here, then, is that we must exercise our reflective judgment in order to have any *purpose* for our actions. The special concept of reflection is purpose, and we legislate the *form* of purpose (subjective, but universal) to our subjective life, including all feeling, by judging in this way. That all actions must be thought of as having a purpose is a requirement of our thinking; in short, it is *necessary*, but in a way that should not trouble pragmatists. Kant says:

> This necessity is of a special kind. It is not theoretical objective necessity, allowing us to cognize a priori that everyone *will feel* this liking for an object I call beautiful. Nor is it a practical objective necessity where, through concepts of a pure rational will that serves freely acting beings as a rule, this liking is a necessary consequence of an objective law and means nothing other than that one absolutely (without any further aim) ought to act in a certain way. Rather, as a necessity that is thought in an aesthetic judgment, it can only be called exemplary, i.e., a necessity of the assent of everyone to a judgment that is regarded as an example of a universal rule *that we are unable to state*. Since an aesthetic judgment is not an objective and cognitive one, this necessity cannot be derived from determinate concepts and hence is not apodeictic. Still less can it be inferred from the universality of experience . . . . For not only would experience hardly furnish a sufficient amount of evidence for this, but a concept of the necessity of these judgments cannot be based on empirical judgments. (C3 237 Pluhar translation)

Kant goes on to argue that this is *hypothetical* necessity (just as Peirce argues), and without it, nothing has a purpose. He carefully argues that we deal with the squirreliness of this ineffability by forming symbols, and that every judgment has a subjective/reflective aspect. There is no avoiding it, we must interpret. In short, Kant refuses to allow subjectivity to be replaced by formalisms, and he holds that judging is situational (particular), fallible, and purposive. It is not mechanical and could never be. It is pre-cognitive, a matter of feeling.

So, without relinquishing necessity, Kant diverts it into a theory of symbol formation and interpretation, and boldly makes the whole of the moral life depend on that theory (C3 sections 58 and 59 –the argument that beauty is the symbol of morality). This tells us why we should try to do our duty even when we cannot know exactly what it is. There would be an aesthetic judgment, shared by everyone, that we were trying to do our duty as we understood it. Such a view opens out onto Royce's ethics of loyalty to loyalty, which I find more satisfying, I must admit.

Updating Kant

Yet, there is value in trying to make Kant into an analytic epistemologist, more as a test of the viability of analytic philosophy than as a measure of Kant. Kant is permanent. Analytic philosophy was a 20th century experiment that became an all-consuming fashion, and is now largely over, in the US, and barren in the other places it is still pursued. It made some contributions to philosophy, but most of them were negative –we now know what happens if we try to turn everything into language (regardless of how broadly, narrowly, or variably language is conceived), and most of those lessons are about dead ends. But they needed to be explored. The narrow and cognitivist interpretations of language have failed, and whatever "intelligence" is, it isn't *just* language, although it does of course *involve* language, when it comes to human beings. The days of denying that animals are "intelli-

[2] I am drawing on the work of Rudolf Makkreel in this summary of reflective judgment in Kant. See his *Imagination and Interpretation in Kant* (Chicago: University of Chicago Press, 1990).

gent" are behind us, fortunately, although the judgment of science and history might run counter to Kant's expectations on that subject. Thus, and this would not surprise Kant, the fundamental question of artificial intelligence must include pre-cognitive and non-linguistic capacities, or it isn't "intelligent" in any defensible sense. Along with some contributors to *Kant and Artificial Intelligence*, I sympathize with the enactivists on this issue.

Thus, I want to bring the perspective of an architectonic Kantian enactivist to these essays, for the benefit of pragmatists, while keeping in mind the purposes of these authors, as much as I can. I very much appreciate the first essay, by Tobias Schlicht, for its comprehensiveness and value as a way of becoming oriented to the spread of questions involved in the central issue of the book. For this reason, I will use its structure as the basis of this assessment. Most of what the book says turn out to be special cases of Schlicht's excellent essay. I will give summaries of the major parts of the book, but most of what I have to say is encapsulated in Schilcht's helpful contribution.

Schlicht's View

I find helpful each of Schlicht's summaries of what AI is, and how Kant has been co-opted or incorporated into the various efforts to *say* what artificial intelligence is. For example, while it is a bridge too far to say that Kant gave a "functionalist" interpretation of mental activity because he was "agnostic" about the substrate (as we saw above, he wasn't so very agnostic), it is not absurd to lift some of the functional discussion from the first *Critique* and use it to work at the question of artificial intelligence. Kant was optimistic about the substrate, and anyone who thinks he was needs to read the second *Critique* –those postulates are not tentative. The very definition of Reason depends on a logical proof (in the Kantian sense of "logic") of the indispensability of Reason for making sense of action, whether in the causal (natural) or the moral (rational) domain. Kant believes we can *have* moral knowledge, even though we must address ourselves and our judgments to the noumenal realm in order to get it (see determinability and determination above). One cannot explain physical action without making this move, at least within the critical philosophy.

But, as Schlicht rightly points out, there are incredibly important functional descriptions of mental activity in the first *Critique*. I do not think most philosophers of mind grasp what Kant means by "Understanding" (merely how the body, below the level of thinking, must sort out the world of sensation, including space and time, by categorizing it) or how it relates to what he means by "Reason," which uses Understanding and depends on it, but does not usually c. onsider the limitations Understanding exercises over (especially) theoretical cognition. That is why a critique of pure reason is *needed* –we can theorize things we cannot experience. And we do. Pragmatists hate that. So does Kant.

But Kant's idea of Understanding maps on to what machines do very closely, and it is indeed the right place in Kant's philosophy to get a full view of the analogy between mental activity and the processing of digital data. The structural features of understanding are indeed a mechanism (a part of the domain of nature) and (according to Kant) best theorized as subsumed under the causal order. Several authors make this point in the book. And for Kant this very *un*free process of Understanding presents problems of a sort closely akin to those argued about in the tradition descending from Turing. Thus, one really must do the work that the functionalists have done, both to extend Kant into the present and future, and also to address the philosophical concerns Kant left to us. Getting Kant "right," in this approach, is not very important, so long as one does not claim that "Kant *is* a functionalist," which would be quite wrong. And of course, Schlicht, while discussing the functionalist aspects, says no such thing, and neither does even Dennett, as far as I know. Rather, Schlicht accurately surveys the scholarship on this issue and expresses the appropriate doubts.

Representations?

Moving beyond classical cognitivism to connectionism, we normally find at the center of the discussion the problem of mental representations, as it has haunted philosophy since Descartes, even as Wittgenstein criticized it (and well before that, it was attacked among pragmatists—what Bowne criticized as "picture thinking" as early as 1882, and James cited his criticisms in *Principles of Psychology*). Schlicht does not summarize this criticism of mental representations, and indeed, it really doesn't come up in the book (and it still finds many able defenders, even a few pragmatists-in-name-only, "pinos"), but it is important to the artificial intelligence debate, so let us pause over it for a moment. The fact that our term "representation" does not exist in German and is conceptually irrelevant to Kant's philosophy has not prevented people from dragging Kant into this argument, on *both* sides of the old dualism. The term "Vorstellung" in German is most closely translated as "presentation," as indeed Pluhar renders it in his translations of the three *Critiques*. There is no "re" in what Kant was describing, no repetition in the mind of what the body experienced, but rather, only an act: presenting, "setting before."

Much confusion has been created by the inability of English speakers to grasp that there is no "re" in the presentations Kant is talking about, and this is not helped by the fact that the German term "Darstellung" is *also* translated as "representation" but has a completely different function and meaning than "Vorstellung" in Kant's philosophy, and actually more closely approaches what many of the 20th century analytic philosophers actually were arguing about. Pluhar translates this term as "exhibition," and that isn't a bad rendering. The difference disappears in English. But if our contemporary epistemologists and philosophers of mind go on the trail of this term in Kant's corpus (and German Kant scholarship), they will quickly realize that it means something quite different to claim that Kant was a "representationalist" than they have imagined, and when the problem is righted, then something quite profoundly anti-Cartesian arises. That change would require them to rework their interpretations around the powers of reflection and purpose, yielding a hermeneutics of the image and imagination (as Rudolf Makkreel has rightly shown in his many writings on Kant).

Presentations come as schemata for Understanding; exhibitions come as symbols. Both are deliverances of our power of imagination in its varied activities –Kant says it does many other things, not just schematizing and symbolizing, but also synthesis, and other activities. This power is mysterious, "an art concealed in the depths of the human soul." (C1 A141/B181) He does not have room for a critique of it, or even sufficient knowledge of it to do so. Kant scholars as a group neglect this power, but it is crucial when it comes to so-called AI, that is, mechanical combining (not to be confused with synthesis, which machines cannot do. Synthesis in general is "the result of imagination, a blind but indispensable power of the soul, without which we should have no knowledge whatsoever, but of which we are scarcely ever conscious." (C1 A78/B103) Imagination runs ahead of sensibility into the supersensible substrate and engages the determin*ability* of that substrate –it is an immediate encounter with possibility. And in that engagement, it does not apply laws of Understanding but rather invents and applies laws of its own. That is what symbolization is for Kant. In the genius, imagination also has a "talent for producing something for which no definite rule can be given." (C3 307) It is creative, original, and (as we saw above) exemplary. It goes without saying, I hope, that mechanized processes cannot do this, and as such, have "no knowledge whatsoever."

The Major Models for Interpreting "AI"

Schlicht goes right past the old problem of mental representations and dives into the newer account of connec-

tionist processing, providing some valuable history and lesser-known passages from Kant that directly address whether Kant might be reconcilable with the multiple realizability interpretations of artificial intelligence. The connectionist thinkers have seen that the power of images is to be the heart of the matter, and in this they are right, I think. But most (who are not specialists in Kant) think Kant has not much to teach them, and in this they are wrong. Schlicht is not as hesitant about this way of tackling the problem. But here I invoke my earlier point about the "spirit" of Kantianism (with a nod to the irony involved). There are many connectionist models, and a sober thinker has no trouble finding sympathy with at least some of them, but I do not think they are Kantian in spirit. The "attitude" of connectionism, as I call it, is disinclined to make use of many Kantian tools, since they think of him as an arch representationalist. That is a worry about the baggage they bring. More about this shortly.

That brings us to the enactivist model. The deep continuity between life and mind is stressed here, and these philosophers rightly land on Kant's discussion in the second division of the third *Critique* of the distinction between self-organizing processes and organized beings. The middle course between brute mechanisms and teleology is created there by Kant and (whether intentionally or not) employed by the enactivists to good effect. This approach enables enactivists to take emotion and feeling into account (and any other precognitive process you might care to mention, including the imaginative processes of which we are "scarcely aware"), recognizing the cognitive value of feeling (a point American thinkers have insisted upon since William James). Here I wish that there had been more authors who support this paradigm in the book, and that Schlicht had stressed the work of Jaak Panksepp, Ralph D. Ellis, Sean Gallagher, and others who hold to the importance of emotion and affect in the development of consciousness. I also wish that enactivists themselves would read Susanne Langer, who was the true empirical pioneer of their viewpoint, and who still has much to contribute, although they don't know it.

I agree with Varela and others that Kant's theory stands in need of revision, that it is under-developed, largely due to the state of the science in his day. But also Varela holds that Kant needs some "naturalizing," and I must agree. The difficulty is that the domain of causal law and the domain of reason are too sharply distinguished from one another in Kant, even if, as we saw, there is a bridge in the determinability and determination of the supersensible substrate (by imagination and perhaps other powers we possess). But that isn't a very appealing solution to pragmatists. Enactivists will see "law" as a metaphor and "reason(s)" as natural, both of which are borne out by our more advanced science. In short, Kant took important steps toward relieving us of the unnecessary dualism of mind and its freedom versus body and its determinism, but did not solve the problem. Enactivists regard the problem as more or less resolved by a better, and more radical empiricism (in James's sense) that does not impose categories on real processes, but derives categories descriptively from patterns we discover in our study of ourselves and the world. But enactivists have, of course, a high bar for what will count as artificial intelligence. Machines will have to feel the world and create their consciousness *from it*. That is very close to what Kant describes as the "feeling of life" in the third *Critique*. (C3 204, 277-278) The recurring controversy about the machine at Google that reported being afraid of death touches on this, but the report falls far short of a human or animal fear. I notice that the contributors to this book allow that Kant would not recognize anything we currently have called "AI" as "intelligence." I agree with them all in this regard.

The turn to predictive processing is a step backward into mechanism from enactivism. Large Language Models and other predictive systems are, if anything, less creative than even parallel processing. It certainly isn't a model for thinking or brain-processes. It is only the metaphor *du jour*. The brain is not some processing unit apart from the

rest of the body, and should not be conceived as a machine of any kind. It is one of the sub-processes in an organized being. Causal relations do not, as far as we know, exist in nature; they are metonymic strategies (taking the part for the whole) for speaking about nature, not provable relations in the world. To claim they exist as such is neither pragmatic nor Kantian –it is beyond our possible experience to observe a determinate cause of an event. The brain seen as a machine is at best an analogy and a deeply misleading one, if taken literally. However, Kant recommended that we press mechanistic explanations as far as they can be pressed, and he really believed in a domain of causes (in a way few scientists do today); he just didn't believe we can experience them directly. The inheritance of Whitehead and many others who showed that causal talk needed to be de-literalized is too often ignored by people who don't seem to understand that contemporary science requires no notion of causation. And this step forward removes science from Kantian philosophy.

Yet, pressing mechanistic explanations, hypothetically, creates a kind of inquiry in which the *analogy* of machine processing (including the predictive kind) to organic processes may be warranted. (C3 417-418) Several authors in this book recognize this fact and employ it properly. If we manage to solve some problems of "learning" in machines on this analogy, then well and good, but it is important to remember that we don't know *why* these techniques work until we drop the crude analogy to a computer and seek deeper and broader analogies (e.g., to non-local quantum processes). We ask: What is learning *in ourselves* (to which we have no complete answer) and what is the machine doing that has structural analogies? One might as well analogize to the way a plot unfolds in a narrative as analogize to mechanical processes, since the former is more illuminating regarding learning. Predictive processing ignores the deeper concerns, such as judging and judgment, which is not mere selection, and which can only be handled by a subtler logic than most philosophers use these days.

For this reason I was very pleased to see Schlict stress the *Jaesche Logic*, and would recommend a balancing with the *Dohna-Wundlachen Logic* of Kant. The failing of most people who study Kant is that they do not recognize that the architectonic rests on a revolution in logic that preceded the critical turn, and which precipitated it. Kant created new categories of inquiry that brought with them several new senses of "necessity," broadening it beyond logic to ontology and epistemology, and providing a basis for a far subtler interpretation of the relation of knowledge and being. People who discuss Kant and speak only of "transcendental necessity," as characterized by "determinate" (apodeictic) judgments, the conditions for the possibility of x, often do not understand "possibility," which may be interpreted either *problematically* or *hypothetically*. As we saw above, reflective judgment is not apodeictic but rather hypothetical. And the 20th and 21st century interpreters do not notice when Kant shifts from one type of necessity to another sense of necessity. This difference makes a great difference, and neither type of necessity applies to what Kant calls "empirical science." This is what Hume contributed to Kant's awakening. Causation (in the sense Kant and Hume believed) cannot be handled by an assertoric logic.

Nativist "AI"

This brings us, at length to the distinction between "nativist AI" and "possibility empiricism," a difference which rests on whether one makes a bright-line distinction between domain specific systems, defining AI strictly in those terms, or allows that AI may be available in non-human forms, i.e., *across* domains. Any time one encounters an a priori and necessary restriction in empirical science, one has encountered a mistake. Thus, adopting the domain-specific stance, as a necessity, *will* block the road of inquiry, eventually. As a special case of possibility empiricism, the domain-specific stance could be used as a hypothesis, but we are not anywhere close to achieving

artificial intelligence with our machines, in the enactivist sense. I do not say it cannot be done in principle, but the quest for a criterion of artificial intelligence is far from over. Every new proposal is more likely to be a false limitation than a goad to better thinking, unless it is taken as a conjecture or suggestion, and never elevated to a principle or a law, as the Turing test effectively was. The various proposals about computer learning are not going to serve us well as criteria for pronouncing some process or machine "artificially intelligent."

Exo-axiology

The underlying idea, as Kant pointed out, is communication. An extra-terrestrial being might be intelligent and we could never discover it unless it can conform to our limitations in communicating, as Kant clearly says (C1 B853).[3] The communicative power takes on forms of limitation that are ontologically non-necessary but empirically devastating to our cognitive capacities. We don't know very much because of the conditions *we* impose on the knowing process (no matter how liberally we define "knowledge"). We would gladly know more, and are demoralized by the fact that we can imagine it but are never likely to know it. Even the least limiting of our cognitive powers, imagination, is still very limited. We must resist the urge to claim that if *we* cannot make sense of something, it therefore doesn't make sense at all. That isn't just Kant's view, it's good pragmatism. Whether something makes sense in the infinitely distant future is for ideally situated inquirers at that time to judge, as Peirce explained and Dewey endorsed. The unknowable, when we make a claim about it, produces antinomies. In the first *Critique* its many erroneous forms are exhibited in the Transcendental Dialectic, the part nobody teaches.

[3] See also Kant's *Anthropology from a Pragmatic Point of View*, trans Günter Zöller (Cambridge: Cambridge University Press, 2011), AP, 237-238; VIII, 215.

The Rest of the Nice Book

The Theoretical division of this new AI book, then, includes various forays into the how and why of Kant's usefulness in addressing AI. I am not optimistic about Richard Evans's effort here, as a piece of Kant interpretation or even as something in the spirit of Kant. He uses the narrow logic of the 20th century and ignores the architectonic. I think he takes himself to be developing a view in the Kantian spirit, but if so, he has not done that. Yet, he has done some important tool-boxing of Kant in a vein that may prove fruitful, and the reason is that we currently do talk to and manipulate our machines mainly using this same narrow logic. That logic is a clunky tool for communicating and is holding us back, as many computer scientists lament. When our communication with the machines, and the cosmic order itself, is more flexible, we will have greater power. But for now, yes, finding formalized solutions to current problems, and theorizing those formalisms in an epistemic way, is potentially valuable.

I am much more optimistic about Sorin Baiasu's effort, since he emphasizes the analogical character of his method of "thinking Kant" in the present. He limits his points of reference to the first three divisions of the first *Critique*, which is too narrow, but he knows he is doing that and is careful not to claim anything about Kant directly. David Chalmers has rightly emphasized the distinction between sensibility and understanding. Now if he would do the same for understanding, reason, judgment, and imagination, with acknowledgement that this is only part of the list of irreducible cognitive powers humans possess, we would have even more progress. What was needed was a generalization of these powers, which is what Whitehead gave us. I have to despair of Chalmers and his followers ever grasping that what they want has already been achieved by people they don't care to read, but we should all welcome the important and true things they assert. It doesn't matter who gets credit, so long

as we find our way to a better account. Yet, I can't help thinking that Whitehead's vocabulary and pluralized formalizations are better than the current language favored by Chalmers and his crew.

Similarly successful is Hyeongjoo Kim's essay answering John McCarthy's call for help in defining AI. One reason Kim is successful is the emphasis on the relationship between transcendental idealism and empirical realism as compatible. They are in fact far more than compatible, and one needs the full architectonic to understand why this is not a problem at all, but connecting the problems of today to the Kant of the first *Critique* in a responsible way is always helpful. More helpful would be to connect Kant's actual philosophy *as a whole* (i.e., the architectonic) to the problems of today, since many are addressed somewhere in that corpus, and tremendous insights are to be had quite beyond the first *Critique*. There is far more to Kant's theory of imagination than its ability to derive a schema that is both intellectual and sensible. Kim's essay makes a very nice transition to the Practical division of the book, and as the co-editor, he is to be praised for the seamlessness of the transition.

With that praise in place, I would have switched the order of the first two chapters in the Practical division, but I will discuss them in the order they occurred. In the Practical division, we shift to what many interpreters take to be a different set of questions, and this is clear in the two chapters regrettably devoted to trolley examples. In fact, the practical and theoretical aspects of Kant's philosophy pose the same questions but in the practical domain; they are asked in a more general way. People often seem to forget that the problem of "pure reason" –why it requires a critique of its own—is that human beings can *think* about things theoretically that they cannot fully *experience*. Yet, since thinking is a kind of experiencing, we have a problem explaining ourselves to ourselves, and besides, we do get knowledge of a sort from theoretical cognition when it is rightly used: synthetic a priori knowledge. The main challenge comes from theoretical cognition, since here we think about things in a way that tempts us to substitute the *results* of thinking for the rest of our experiencing.

As with Tobias Schlicht in the theoretical division, Dieter Schönecker (also a co-editor) makes a number of the points I would make in his chapter. That temptation to think beyond the bounds of possible experience comes in two forms: we think things are true of the world that are really only true of our thinking (which creates dialectic); *and* we think we can do things we cannot in fact do (which brings about moral error). Both problems take the form of determinate judgments gone awry. These are the problems of pure reason and pure practical reason. Since thinking is primarily practical, and since theoretical cognition has its basis in practical action, we have then, pure practical reason, which tempts us only to think about things we cannot do, and full-fledged theoretical cognition, pure reason, which tempts us simply to make inferences that fail to conform to the limitations of our power of determinate judgment.

Yet, when the problems associated with our powers of pure practical reason and theoretical cognition have been addressed, there remains the more important problem of the genuine practical relation between our practical thinking and our practical action, and how we make sense of how each comes from the other. This is where so many interpreters get lost in the theory of the form of the will, in the *Groundwork* and in *The Metaphysics of Morals*. These are books about how we ought to *think* about the will, the norms of pure practical reason, not about what in particular cases the will *should will*, which is practical and situational. The issue of "agency" and the associated problems of self-determination and autonomy, come from the logical extension of our powers of action in a *problematic* rather than a hypothetical logic. This is not about what we might do, but only about what we *can* do, and it definitely isn't about how we ought to think about what we might do. This is the reason that determinate judgment is the relevant logic:

we are concerned with how what is actual relates to what is possible, i.e., problematic logic.

This kind of logic can also be extended to an infinity of actions which we *may* perform (not might –ought implies can), and here the aim is to know what we should and shouldn't do, such that knowledge is the aid of action. Here the stakes are higher, since we only risk *logical* error when we misuse theoretical cognition. No one dies of a bad theory unless it is coupled with further action. Failing to know what we ought to do is far more dangerous than failing to know how to think about what we may do. Yet, knowing what we are doing (and may do) is very important. For this reason we cannot wholly divorce our agency from our knowledge. If we do, the moral quality of our actions disappears and we are counting on luck. That is a fair description of the amoral quality of mechanical thinking. As our authors point out, then, in the practical section of this book, agency requires a robust autonomy. That is the basic reason they are skeptical of the likelihood we can ever create Kantian "AI."

Lisa Benossi and Sven Bernecker are certainly right to make autonomy and freedom the center of their case, but they take the view that robots "could not possibly" be moral agents, to which I must reply that this assertion isn't knowable. Although they also say that this is "unlikely" and that it is a "low probability," rather than "impossible" later, I think they mean what their abstract says, which comes to "never." Otherwise, they don't need the logic they use, which isn't probabilistic. The negation of robot moral agency has both a logical and a practical dimension. Logically speaking, they use Harry Frankfurt's much narrower logic rather than Kant's, and that makes their case irrelevant to Kant, which they do not acknowledge. It is at most Kant-like or in the spirit of Kant, but I deny the latter. Kant doesn't think like they do. His own logic is perfectly adequate for addressing this question. We don't need Frankfurt. So I am less enthusiastic about this approach, even if, on its own terms, there is no serious problem. It isn't a sin to depart from Kant.

For Kant, and his ethics, saying "never" requires that we prohibit a possibility (as they say "could not possibly," rather than "can not possibly") and that may be done in one of two ways: problematically or hypothetically. The first implies that no finite series of steps leads to moral agency for robots, while the second requires that no possibility of robot moral agency is consistent with any actuality. Both arguments can be made, but neither can be knowledge. Both theses are under limitations Kant explains in the Transcendental Dialectic of the first *Critique*.

For the same sorts of reasons Benossi and Bernecker give for saying that we cannot conclusively answer the questions about whether we *should* seek to create machines that have moral standing or even personhood, they should also say that we cannot know whether our creations can be moral agents. After all, we create new moral agents every day, in having children, at a rather alarming rate, and often with the aid of machines, while others we keep alive by means of machines, and others are part machine. There is no bright line distinction between "artifacts" and ourselves, and their conditions, that "a robot uses sensors to detect aspects of the environment, software to reason about it, and actuators to interact with it," (pp. 147-148) will have to be adapted or discarded eventually. In the future our integration with the machine world will increase rather than decrease, and so the practical changes must drive the theoretical positions we take, not vice-versa. Is it permissible to prohibit making these machines? Yes, I think they are right. That prohibition is permissible. But it doesn't matter. When we *can* make machines that deserve moral standing, we *will* make them. We might as well think about what we are going to say about their agency and standing now, since their existence is a matter of when, not whether. That is pragmatism, in my opinion.

Schönecker's case that machines cannot have practical reason because it is free, and they are not, is closely in keeping with what I would expect Kant himself to say. It is

clearly Kantian in spirit, but I think both Schönecker *and Kant* overstate the matter. This is a pragmatic question as well. The separation between the domain of causal laws and the domain of reason and free action is too stark in Kant's first two *Critiques*. It thus serves us badly when taken apart from the full architectonic. When we include the reflective power of judgment and its critique, our thinking about freedom is broadened beyond the way that reason is made determinate in understanding and in action. There is a kind of judgment that is reflective and draws on the special concept of purpose. This form of judgment is exemplified in the judgment that "x is beautiful," based on a "free play" of understanding and imagination. We saw this at the beginning of this essay. This form of judgment then raises the question of purpose in nature, and the answer is a hypothetical proposition: it is *as if* nature were made for our purposes (including our self-improvement for an everlasting time), and for our freedom of action. Indeed, we cannot help thinking so, Kant says. (C3 434-436)

This feature of nature, which Emerson called "the sentiment of virtue," leaves us with the experience that all of nature is working for our good, or at least our betterment. Thus, Schönecker's claim that to be free is to be able to will the good seems right, but it needs context, and the context is the domain of nature. The determinate nature of our understanding and our practical reason, and the concomitant form of determinate judgment, is not sufficient to undergird our experience of freely moving our bodies, as Kant points out, so there must be more. I take Schönecker's final analogy about swimming and flying (pp. 185-186) to capture this feeling, perhaps better than his earlier argument does. The question is what can we know, do, and hope regarding purpose? That is the real role of practical reason in Kant, not just acting freely and autonomously legislating our duty to ourselves. We want more. We want nature to cooperate in our purposes.

I suggest that Schönecker has focused too narrowly on practical reason in its determinate form. Ironically, the higher bar for freedom, the freedom of thinking, imagining, and even acting as organized beings in a world of purposes, is more attainable for robots than the lower bar set by Schönecker's focus on practical reason. It is easily imaginable, without falling into dialectic, that machines might have purposiveness *as* self-organizing processes –this is what many interpreters mistake for "learning." If that is so, then the line between self-organizing processes and organized beings is thin. We want a principle for thinking about *life*, and if self-organizing processes are not life, they are very close to it. (C3 422-429)

Elke Elisabeth Schmidt unfortunately goes the way of the trolley example and "conceptual analysis" for the sake of tapping or pumping or refining or adjusting our "intuitions" –not in the Kantian sense. She does not state this, but the outcome of conceptual analysis is the forming of moral intuitions. One does not need such examples to form Kantian moral judgments. None of this, then, has anything directly to do with Kant and I have yet to see a genuinely Kantian response to the so-called trolley problem. The genuine response is that the problem itself is clearly just the sort of thing avoided by the limitations on theoretical cognition set out in the Transcendental Analytic and the mistakes illustrated in the Transcendental Dialectic. The relevant "programmer" isn't the one who sets the trolley car programming for autonomous trolleys, it is the philosopher who poses the problem and then insists it has some importance for our thinking. It does not.

Conceptual analysis is foreign to Kantian philosophy, as indeed it is to pragmatism, and our moral intuitions are not guides to anything reliable, no matter how they are pumped or refined or adjusted. As Kant says, don't trust your natural inclinations as moral guides. Here many pragmatists will protest, but I remind them that education and imaginative rehearsal are methods of intelligizing practice, and they are morally required for the melioration of any situation, since all situations carry valuative features. Kant is saying nothing different when

he cautions us from following our natural inclinations. He is saying, in effect, "don't imagine that following your uneducated desire will lead you in the better direction." Pragmatists hold the same view. This kind of philosophy, driven by conceptual analysis, is Thomson, and Foot, not Kantian in spirit or letter, and not really even drawing much from the Kantian toolbox. Thinking is what is required by Kant, not analogizing without the aid of genuine practical reason. Kant confines intuitions to their determinate forms, or, under the right circumstances, their reflective forms. Moral imagination is not disciplined by trolley examples, and Schmidt's case is an example of trying to cut Kant's moral philosophy off from the architectonic. Kant would not expect good results from that and neither do I.

Ava Thomas Wright takes for granted, as I do, that morally autonomous machines will be built as soon as they can be. Unfortunately, she falls victim to the same problems Schmidt has. The problem is how do we handle issues that will result from conflicts with these machines, whether they are driving machines or trolleys, or house cleaners and factory workers. Perhaps we got a foretaste of this kind of conflict when the Google employee refused to turn off a machine because it reported being afraid of death, but this case was not a thought experiment, such as Schmidt and Wright apparently believe are the relevant analogies to life and death decision-making. But those self-driving cars are motoring their way into our moral world. The problem isn't hypothetical anymore. If they lack agency, we will still sue the corporation that kills our loved one with one.

Wright seeks to set these conflicts on firm Kantian ground by placing them in the domain of right rather than virtue. This is worth thinking about, but it also presupposes an interpretation of Kantian moral thought that separates it from the architectonic. Keeping Kant's moral philosophy within the architectonic provides greater interpretive freedom and does not oblige us always and everywhere to place such conflicts wholly within one side or another of a determinate conceptual difference. It also keeps the moral philosophy in proximity to pragmatism. The question of purpose is the most comprehensive question in Kantian moral philosophy, and it is reflective in form rather than determinate. I recommend a broader analysis. We will find that things Wright thinks are necessary are not quite *as* necessary.

The longest trolley ride into things not relevant to Kant is relieved by Claus Dierksmeier's chapter, which does acknowledge the importance of the architectonic. His emphasis on purposes keeps his argument well within the tradition of Kantian moral philosophy as Kant pursued it himself. The insight about partners rather than parts strikes me as very much in keeping with the organicity of Kant's analogy between nature and society. Working the analogy as a "symbol," in Kant's sense, would be even better. A symbol, for Kant, is a complex but unfinished product of our imaginations, in which what is beautiful stands for (without replacing) what is moral, and hence what seems purposive in nature can serve as a guide to what is purposive in ourselves, as natural beings. This organicity applies especially to our social organization, which grows from our *sensus communus*, which is natural in us for Kant. I find much to praise and nothing to criticize in this essay. I only recommend that Dierksmieier consider his view in relation to the Kantian symbol. Kant is not a deontologist, in the sense employed by the moral philosophers of the last 100 years. He is a commonsense moral philosopher, owing more to the Scots and to Rousseau than to any strict formalist version of moral thought. Dierksmeier is trending this direction, but could perhaps benefit from some of Cassirer's essays on Kant's relation to Rousseau.

That brings us to the final chapter, the sole essay in the division of Aesthetics. I do not agree with Larissa Berger that Kant's requiring of disinterestedness for a judgment that "x is beautiful" is phenomenal, although I agree that it has *implications* for the phenomenal. The distinction between phenomenal and noumenal, or, in first *Critique*

language, of appearances versus things in themselves, is brought together in the third *Critique* in the discussion of the supersensible substrate. We saw that discussion in the part above, in the excerpt on determinability and determination of that substrate. But there is more (see C3 176, 196, etc.) Somehow Berger manages to write a huge number of pages about the phenomenal in Kant as it applies to pleasure and disinterestedness without once mentioning the supersensible substrate in the third *Critique*, which cannot be theoretically cognized, but it can be felt, and indeed must be appealed to in order to solve the Antinomy of Taste (C3 339-344).

This oversight is unfortunate because it disqualifies her claim that "quite strikingly, Kant seems to be confident that he can just presuppose [the] T[hesis of] D[isinterestedness] as a brute matter of fact." (p. 270) This is far from being accurate. I recommend that Berger consider the discussion of whether there is an ontological unity of the noumenal and the things in themselves in a felt supersensible substrate before she concludes that the disinterestedness requirement is presupposed or ungrounded. Before she offers her three speculations as to why Kant does not ground his account of disinterestedness further (p. 272), she might consider her own point: "Unlike Nagel, Kant does not use the term 'subjective' to refer to the phenomenal character of experience." (p. 279) What, then, is "phenomenal" for Kant, since subjectivity is noumenal, as indeed is personhood, at least when it comes to our standing as "persons." The other person is the only truly sublime experience we have, so it seems to me that the Analytic of the Sublime bears on the case of disinterestedness, since we can never judge the beauty of a *person* disinterestedly. The limit case helps us define the ground of disinterestedness.

The fact that the term "noumenal" never occurs in this chapter by Berger suggests to me that Berger has not really considered the meaning of the word "phenomenal" in Kant, which is the contrast term to noumenal. The realm of appearances in the first *Critique*, and as determined by our power of understanding, is broadened to the domain of reason in the second *Critique*. Both are related by reflective judgment, as we saw at the beginning of this review. There is a realm of things that can be thought about and acted upon, which is phenomenal, and thought about but not acted upon practically, which is noumenal. The noumenal is the ground of the phenomenal in the domain of reason, which is to say that we cannot make sense of our own actions unless there is a wider realm of meaning for our actions than we can determine by acting. Without imagination's mysterious power to determine the determinable in the supersensible substrate "we should have no knowledge whatsoever." In short, there must be a world we never made with meanings we will never fully know. The issue of the supersensible substrate in the third *Critique* is the question of whether the purposes we feel in acting might bring the purposiveness of nature to our power of reflective determination of ourselves as subjects. The discussion of the supersensible substrate in the third *Critique* is, indeed, inconclusive, but it introduces the problem of natural teleology, which may be said to offer some solutions.

In all, the volume is a nice contribution to current discussions, and while there is not much Kant *scholarship* in it, there is a good bit of thinking in a Kantian spirit, which may be added to the broader discussions of AI that are going on. Pragmatists could do themselves a favor and pay attention.