Table of Contents

Portfolio Introduction

The way that most people conceptualize the external world, objects in the physical world exist wholly independent of any subject's perception of them. While intuitive, this notion of the world is very limiting. Taken to its conclusion, it entails that we cannot know anything more about cause and effect than simple probability. With this view of the external world, even our most scientifically rigorous predictions are no more than mere habit and superstition. Current machines employ this world view, so their interactions with objects are entirely based on probability and patterns. For example, current machine intelligence uses patterns to recognize images; however, they are not capable of interacting with those objects in the external world.

The next step for human society is the creation of artificial intelligence capable of reasoning. Past and current machines have not needed to use reason; they perform menial tasks or work with probability. Reasoning machines, on the other hand, can causally interact with objects around them. This means that they can recognize and work with underlying structures instead of simple patterns. This ability to causally interact allows machines to complete tasks which only humans previously could. This type of artificial intelligence could utilize the focused precision and accuracy of a machine to complete those tasks even more efficiently, powerfully expanding the human capacity to make predictions. In this way, the creation of reasoning machines will shape the future of human life.

In this portfolio, I use the philosophy of Immanuel Kant to isolate the elements of human cognition necessary for causal reasoning and clarify how these components can be applied to the creation of reasoning machines. To accomplish this, I first examine why Humean reasoning currently limits artificial intelligence. An understanding of David Hume's problem of induction highlights the inefficiency of correlation-based reasoning currently used in machine intelligence and illustrates the qualities required to go beyond probability. Next, I outline Kantian transcendental idealism and its account of causal reasoning. Through an explanation of the cognitive faculties surrounding experience, this theory of knowledge determines the elements of cause and effect. Finally, I show how applying Kant's account of mental activity can assist in the construction of an artificial intelligence that could employ reason even in new and increasingly abstract situations. A reasoning machine capable of replicating human cognition in this way would overcome the boundaries which have limited artificial intelligence so far.

Piece 1: Hume's Problem of Induction and Modern Machine Intelligence

1.1 Introduction

David Hume's problem of induction highlights the difficulty of conceptualizing objects in the external world as wholly independent of subjects. The problem concludes that we can never determine cause and effect using our senses (inductive reasoning). This conclusion reduces cause and effect to probability calculations. Correlation-based reasoning is what limits modern machines from causally interacting with the world around. Thus, finding the qualities of a theory of knowledge which could resolve the problem of induction is the first step toward reasoning machines.

1.2 Humean Inductive Skepticism

Scottish empiricist David Hume (1711-1776) posed a very significant challenge to the validity of reasoning about causality through inductive inference with his problem of induction, and a resolution of this problem lends insight into the conditions of cognition that would be necessary to reason about causality. Causality is a relationship between events where one necessarily leads to the other. Since many of our beliefs are based on inductive inference, we should want some sort of justification of induction's validity, especially in determining causality.

Hume argues, neither deduction nor induction itself can be used to justify inductive inference. Deduction cannot justify induction for two related reasons. The first reason is that deduction deals with necessary knowledge, propositions that must be true. In contrast, induction also deals with contingent knowledge, propositions that are true in some possible worlds but false in others. The second reason is that deduction is non-ampliative. This means that it does not involve new information being inferred as it deals with already known information (Morrison 445). On the other hand, induction is ampliative; it takes current information and applies new information to create new knowledge. Additionally, induction itself cannot justify inductive inference either. If we argue that our inductive inferences in the past have proven true, and thus induction is valid, this presents two more problems. First, this argument is circular as it uses induction to prove that induction is valid. Second, this begs the question of whether our inductive inferences have been true in the past. A person skeptical of inductive inference, such as Hume, would likely deny that premise. For these reasons, David Hume concluded that inductive inference can never be valid.

1.3 Humean Causation

Hume's skeptical conclusion has numerous implications on traditional modes of thought. If true, Hume's conclusion subverts the basis of modern science itself as inductive inference makes up the foundation of the scientific method. Induction is the observation that begins the scientific method, and inference leads to the predictions that define scientific hypotheses. Underlying this thought is the notion that what we will experience in the future will conform to what we have experienced in the past. Hume argues that we cannot find necessary connections between things we call causes and effects; rather, we expect that certain things will happen based on habit. If I see a ball collide with the floor and hear a sound in conjunction with that event enough times, I will expect to hear a sound every time two things collide, but Hume posits that my expectation in these instances is simple habit, not a determination of causal relationship. Thus, in Hume's view, causation is no different than correlation, and causal reasoning becomes mere probability.

Trying to use Humean causal reasoning has serious limitations. Since cause and effect rest in correlation and probability, most actions can only come through a great deal of trial and error. This situation partially arises because Hume's inductive skepticism goes hand in hand with his empiricism. Since empiricism asserts that causal reasoning comes only from trial and error, there can be no sweeping causal claims applied to underlying structures in any experiences. This leaves no room for learning about the qualities or nature of relationships or objects. Thus, working through problems is slower and less efficient than necessary.

1.4 Modern Machine Intelligence

Current work on robots, artificial intelligence, and machine learning has thus far employed a Humean sort of causation because most researchers working on machine vision have "focused largely on recognition, rather than understanding structure" (Sloman 47). The limitations of this form of causal reasoning limit artificial intelligence as well. Machines exhibiting Humean causation focused on correlation and probability often can accomplish tasks, but they cannot understand or even know how they accomplish things or what would have happened if they acted differently (Sloman 48). This mode of cognition possessed by modern intelligent machines lacks truly causal reasoning, so they cannot progress beyond the basic aspects of behavior and intelligence.

1.5 Resolving the Problem of Induction

In order for an artificial intelligence to possess causal reasoning, the structure of its cognition must be built in accordance with a theory of knowledge that can resolve the Humean problem of induction. To accomplish this, the theory of knowledge must satisfy two criteria: it must be anti-realist and it must account for causal understanding of the relationships between observed objects.

The first criterion, anti-realism, is initially counter-intuitive. Epistemological realism regarding the physical world asserts that objects in the physical world exist, and they exist wholly independent of any subject's perception of them. Common sense easily accepts this view;

however, our senses could never truly represent all the qualities of objects in the world. For example, we know that qualities such as color are not truly intrinsic to objects, as they are very specific reflections and refractions of light. Furthermore, human eyes are not capable of perceiving many wavelengths of electro-magnetic waves. Since these qualities are merely quirks of human perception, we can never be certain of our knowledge of these independently existing objects. The only way to even verify the accuracy of our perceptions would be to somehow step outside of human perception and compare our representations of objects to the object's inherent qualities. This, of course, is impossible because humans are always trapped behind the veil of their own perceptions. As scholar Richard Schantz sums up, "[t]he totality of our beliefs about the world is one thing, the objective world quite another" (Schantz 478). In contrast, anti-realism holds that the only philosophically significant objects in the context of the external world are the ones created by our senses. In this view, our representations of objects create a distinct, ontologically weaker world. If the inadequacy of realism is that the human mode of cognition does not necessarily conform to objects, anti-realism's solution is that all philosophically significant objects are constructed by human cognition and thus conform to it. In the context of machines, those representations of objects translate as sensory data.

The second criterion, accounting for causal understanding of the observed relationships between objects, deals with the structure of cognition. Hume's skeptical argument concludes that the senses are unable to determine causality in the relationships between objects. For Hume, constant conjunction of events creates habits of expectations within the observer, and no knowledge beyond that can be gained. Thus, a theory resolving skepticism must be able to account for the senses' ability to determine causal relationships and interaction between objects.

Having determined the two criteria for a theory of knowledge which can solve Hume's problem of induction, we now have the starting point for the creation of reasoning machines.

Piece 2: Kantian Transcendental Idealism As a Model For Consciousness

2.1 Introduction

In searching for a theory of knowledge that can resolve Hume's argument, the clearest starting point is the work of Immanuel Kant. In fact, Kant's writings on mental activity and cognition react specifically to David Hume. He states this quite explicitly in his *Prolegomena to Any Future Metaphysics*: "[T]he remembrance of David Hume was the very thing that many years ago first interrupted my dogmatic slumber and gave a completely different direction to my researches in the field of speculative philosophy" (Kant 10). It is for this reason that Kant's philosophy forms a very complex, yet coherent, response to Hume's problem of induction. Thus, Kant's theory of knowledge and cognition, known as transcendental idealism, provides a very strong basis for causal reasoning. This type of causality, as well as Kant's formulation of the self, has important implications in the realm of artificial intelligence, but the explanation of these implications requires an understanding of Kant's complex philosophy of transcendental idealism. Therefore, this piece of the portfolio is dedicated to detailing the structure of this theory.

2.2 The Copernican Strategy

To introduce transcendental idealism, Kant makes an analogy to the argument of Nicolaus Copernicus, the Polish astronomer and mathematician. Copernicus first established the idea of heliocentrism, which holds that the apparent movement of the sun is an aspect of the subjective condition of an observer on earth. Because the observer is on the earth, which is moving around the sun, it appears to the observer that the sun itself moves. Analogously, Kant's theory holds that the subject of knowledge actively conforms perceived objects to its own mode of cognition. This, very straightforwardly, appeals to anti-realism, satisfying the first criterion for resolving skepticism. As Kant scholar Sebastian Gardner summarizes, this Copernican revolution rests on the idea that "reality in the weaker sense is something that we can know precisely because we constitute it" (Gardner 41). This concedes to the inductive skeptic that knowledge of external independent objects, known as "things in themselves," cannot be known. However, it holds that objects can be known by their appearance.

2.3 Space and Time As A Priori Intuitions

The next major aspect of transcendental idealism involves conceptualizing spatial and temporal relationships as given aspects of human cognition. This section will require a few clarifications of terms. "A priori" translates from Latin roughly as *from before* and describes claims of knowledge justified outside of experience. In contrast, "a posteriori" translates as *from after* and describes claims of knowledge justified by experience. Kant refers to "intuitions" as the representations and perceptions by which objects are presented to the subject through experience. Intuitions are distinct from "concepts" which are the representations by which subjects can think about objects. Kant elaborates in the *Critique of Pure Reason* six total reasons for why space must be an a priori intuition.

The first two arguments explain why a subject understands space prior to experience. First, space must be a priori because it is presupposed for outer experience to even be possible. Outer experiences, as used by Kant, are experiences that are distinct from oneself. In order for any sort of experience to be categorized as outside of oneself, the mind must be able to conceive of the spatial relationship between the subject and the event being experienced. Second, space is a necessary condition for the possibility of experiences in the first place. If the mind did not conceive of spatial relationships, events and experiences would have no place to occur; experiences are only possible as they can be perceived (A23-24/B38-39).

Having demonstrated that space must be understood before experience, Kant's third and fourth argument show that space must be an intuition. The third argument asserts that space is a singular entity. When speaking of different spaces, they are always thought of as individual parts of one large singular space. These smaller parts are conceived of as being inside that larger, unitary space. The conception of that unitary space necessarily precedes the classification of smaller parts of space because those smaller parts are only identified by their relationship to space as a whole. Since space represents one individual object, it must be an intuition because the representations of concepts, the representations by which subjects can think about objects, always require some other independent representations, as well. The fourth argument shows that space is an infinite idea. Space has no distinctly represented boundaries and is infinitely divisible, so it contains an infinite number of potential parts and separations. A finite mind cannot conceptually grasp this infinite magnitude, so space must be an intuition (A25/B39-40).

The fifth argument uses the idea of incongruent counterparts. Kant makes this idea clearest with the example of left and right hands. If we define those objects in space based solely on the spatial relationships of their components, the left and right hands are identical. Each of their components has the exact same relationship of parts and placement. However, the left and right hand are obviously not identical. Their difference is irreducible to the conceptual markers of how parts relate to each other, so the notion of incongruent counterparts only makes sense if space is taken as intuition.

The sixth argument deals with geometrical judgements as synthetic a priori. This requires another clarification of terms. Kant uses the word synthetic to mean the opposite of the word analytic. An analytic judgement is a judgement where the predicate is contained within the idea of the subject, e.g., the statement "all triangles are three-sided." It falls within the identity of the idea “triangle” that it is a three-sided object. These kinds of judgements add no new knowledge. Synthetic judgements are the opposite; the predicate of the judgement is not contained within the idea of the subject. These judgements are not based on identity or definition, and they synthesize independent information, e.g., the statement “the teacher is calm.” The emotional state of the teacher is not contained within its identity, but this synthetic judgement combines information and ideas to create new knowledge. Kant argues that Euclidian geometrical judgements, aside from definitions, are synthetic. Kant here deals with Euclidian geometry and three-dimensional conceptions of space because these are the ideas bound to the way humans normally perceive the world. These judgements are also not solely based on experience, and they are formulated on the concepts of spatial relationships. Therefore, space must be a priori.

Kant’s arguments about time follow similar patterns, but time deals more importantly with inner experience. Time is a presupposition for those inner experiences, and it is also a necessary condition of experience as temporal relationships allow for the ordering of representations.

The determination of space and time as a priori intuitions is important for the theory of knowledge because these intuitions underlie all spatial and temporal concepts including the representations of objects perceived within space. If these ideas are a priori intuitions, then they are pre-structured within the human mode of cognition, and this structure is a necessary condition for experience.

2.4 Faculties of Cognition

To deal with the information found in experience, Kant outlines four relevant cognitive faculties. The first cognitive faculty is “sensibility.” This faculty deals specifically with intuitions. Intuitions are presented to the subject in what Kant calls a “manifold,” defined as a multiplicity of sensations. In the context of machines, this is effectively the raw sensory data. The second cognitive faculty is “understanding,” which addresses concepts, or the representations by which subjects are able to think about objects. These concepts must be applied to intuitions. (Gardner 66-67). Kant also points out the necessity of the combination of these two faculties, as concepts or intuitions on their own do not produce knowledge and are empty (A51/B75).

The third cognitive faculty to discuss is “apprehension.” Sensory data within the manifold must have unity to be recognized. An example helps simplify this concept. If one perceives many data points of red colors, apprehension combines those senses as a unified patch of red light. Apprehension enables recognition of the combination of data in the sensory manifold. This conjunction cannot come through the senses and experience, for even if all the data are combined, the act of recognizing that combination is distinct from just perceiving it. Thus, this kind of synthetic unification must be a priori (Gardner 127-128). The fourth cognitive faculty of interest is “imagination.” This faculty can represent objects even when they are not present in the senses through the manifold of intuition. Imagination reproduces objects and fits them within the a priori structures of consciousness.

2.5 Transcendental Unity of Apperception

The transcendental unity of apperception lies at the center of Kant’s theory of knowledge. In his *Critique of Pure Reason*, Kant explains that the “transcendental unity of apperception turns all possible appearances that can come together in one experience into a connection of all these representations according to laws” (A108). The transcendental unity of apperception is the unification of all experiences based on the single factor that remains constant among them. That single factor is the subject of perception itself, the “I.” This allows for the connection of new experience to previous experience. Kant argues that unifying all those representations as belonging to a single subject, in one self-consciousness, allows for the synthesis of those representations. For representations to qualify as "mine," they must be attributable to a single subject, and objects must be perceived in the context of a thinking subject's senses. Synthesizing representations of intuitions and representations of concepts creates knowledge.

With this idea of the transcendental unity of apperception, knowledge of the self is not something given through the senses. The self could only be sensibly known through perceptions from inside the body. However, these perceptions are empirical temporal objects which vary over time, so those ideas of the self would not be constant. Since the self cannot be known through experience or the senses; self-consciousness must be a priori. Sebastian Gardner summarizes this transcendental apperception as "consciousness of thinking" (Gardner 148). This idea which comes from intuition does not constitute knowledge of the self, but it provides the ground of representation of the self as spontaneous. Spontaneity is the relation between the "I" as the subject and its cognitive actions which allows the subject to label itself as an intelligence. The most important idea to keep in mind from the transcendental unity of apperception is that the unification representations is a necessary condition of experience and of synthesizing knowledge.

2.6 The Analogies of Experience

To explain the way human cognition accounts for different aspects of experience, Kant presents three analogies of experience. Within these analogies, Kant provides an explanation of how the senses can determine causality by finding a necessary connection between the succession of events.

The first analogy of experience deals with the permanence of substances. When a person puts a pan on a stove, some qualities of the pan change, but the pan still stays a pan. Qualities of the pan are changing, and the pan is undergoing alterations, but the pan itself is considered to be the same substance. Kant explains that all objects' appearances contain the permanent substance of the object, and the changing qualities of the object are its different modes of existence over time (A182). In the context of time, all appearances are either coexistent or in succession, and change affects appearances over time. In this analogy, the ideas of change and alteration differ. Change occurs when some quality disappears and is replaced by something else. Alteration occurs when a change happens in the qualities of an object. As a result, when an object undergoes an alteration, it does not disappear completely and get replaced. Kant argues that since humans can identify singular objects that exist through their variable appearances in time, there must be something unchanging that is representative of time, which is a framework in the structure of consciousness as a priori intuition, in general. He thus concludes that all changes and successions of appearances are just alterations of an unchanging substance. Since Kant derives this by looking at experience and then working backward to determine the necessary condition for the possibility of that experience, this conclusion is needed for experience and cognition to occur.

The second analogy of experience deals with causality. It begins with the notion that every event has a cause. While the first analogy established that change is an alteration in substance, the second analogy addresses how a subject can know whether change comes from an object or whether change is just the subject's differing perception. Causality can be determined in part by whether change occurs as a facet of the object, which is objective change. Since all change occurs as different successions of appearances within time, for change to be objective, the relation of successive events must be necessary and irreversible. Kant provides exemplary illustrations in sections A190-3/B235-8 of the *Critique of Pure Reason*. In the first example, a subject walks around a house in a counterclockwise direction. In this case, the succession of different appearances is determined entirely by the subject since the order of succession could have been exactly reversed if the subject had walked around the house clockwise. The subject would still be observing the same object or event, but the changes in appearances would be different. Since the order of successive appearances is neither necessary nor reversible, this change is subjective. The second example given by Kant describes a subject observing a ship moving downstream. In this situation, the order of successive appearances could not have been reversed, for then the subject would be observing a different event, a ship moving upstream. This change is objective. This kind of determination justifies a conception of causality as a necessary relationship between events because an a priori rule allows the subject to find necessary and irreversible change in objects. Thus, this analogy proves that the senses are capable of accounting for causality.

The third analogy of experience focuses on coexistence and the causal interactions between objects. Objects or events can be coexistent, which mean they have the same location in time, when a subject can view the objects or events in a reversible order of succession. If a subject can witness event A before event B but can also possibly view event B before event A, the subject can know for certain that event A does not cause event B and that event B does not cause event A. If either event caused the other, then the cause would occur before the effect every time. This is because cause can be represented as the conditional judgement "if event A occurs, then event B will occur."

The analogies of experience establish the aspects of cognition necessary for experience and for inferring causality. Thus, Kant's theory of knowledge and mental activity succeeds in overcoming Hume’s inductive skepticism and succeeds in establishing the conditions for a cognition that can progress past Humean causation.

Piece 3: Toward Reasoning Machines

3.1 Introduction

Immanuel Kant’s transcendental idealism defines the cognitive faculties necessary for reproducing intuitions and concepts to fit within the a priori structures of consciousness, allowing subjects to interact with all their previous knowledge. Additionally, Kant’s categories of judgements and his notion of spatio-temporal relationships enable the behavior of reasoning machines to be traced, providing controls that would safeguard against choices that are heinously unethical or could endanger human lives. In these ways Kant’s account of mental activity can assist in the construction of reasoning machines that can work even in novel and increasingly abstract situations.

3.2 Using Transcendental Idealism to Structure Cognition

In his writings about his theory of knowledge, Kant often employs a regressive argument structure. This means that he takes certain premises and ideas and then works backward from those ideas to determine the conditions necessary for the existence of those ideas. This argument structure proves very important in the context of cognition because Kant employs it in formulating the analogies of experience. Especially in the first analogy, he works backward from experience itself to determine the structure of substances necessary for the possibility of that kind of experience. This type of argument and determination of necessary conditions appears throughout Kant’s justification of space and time as a priori intuitions, the cognitive faculties of sensibility and understanding, and the transcendental unity of apperception. As such, this theory of knowledge lends insight into several parts of mental structure that are necessary conditions for perception and cognition and provides an optimal path for fully replicating human cognition in intelligent machines.

The primary place where this theory of knowledge must be applied in structuring artificial consciousness is in the synthesis of intuitions and understanding. Kant's notion of intuition is best represented in terms of machine function as raw sensory data. Modern machines are very good with the cognitive faculty of sensibility. As discussed in piece 1, current work in machine intelligence is largely focused on visual object recognition, and this field has advanced greatly, so machines nearly have the cognitive faculty of apprehension in that they can combine data points into unified objects. The next step in machine perception is to form the faculty of understanding. Progress in this field can also be seen in machine intelligence as the recognition of objects requires concepts to be applied. However, the current capability of recognition deals less with applying abstracted concepts and more with sorting patterns into categories. This is also why many machines can be tricked to classify images into categories they most definitely do not fall under. For example, LabSix, a research group composed of MIT students, tricked Google's InceptionV3 image classifier into thinking that a 3D-printed turtle was a rifle. The group used an algorithm to "generate both 2D printouts and 3D models that fool a standard neural network at any angle" (Athalye et al.). The reason this is important is not just because machines might misclassify objects. This research shows that even the most intelligent modern machines do not actually have knowledge about the qualities of objects; they attempt to recognize patterns. Only through the synthesis of representations from sensibility and representations from understanding can true knowledge be formed. Updating this language to a more modern form, author J.D. Casten, in the book *Cybernetic Revelation*, describes this conception of knowledge as a "semantic network," a web of "different types of concepts with various connection" (Casten 269). To contribute to this semantic network, the faculty of imagination would also need development, as that web of concepts with connections requires reproducing intuitions and concepts to fit within the a priori structures of consciousness.

Underlying the faculties that enable understanding are the aspects of Kant's mental theory that discuss unity of the self. Perception relies on unifying intuitions and sensory data into objects, but that object unification requires a unitary subject. This unification occurs with the transcendental unity of apperception. In identifying the self as the subject of all common perceptions, concepts from previous experience can be connected in the semantic network. Through combining the faculty of imagination and the transcendental unity of apperception, Casten explains that we "may imagine that the implicit knowledge of the understanding that shapes our inner sense becomes explicit and hence workable by the imagination through constructing and connecting the similarities found in the plurality of experiences" (Casten 286). This analysis indicates that to construct any cognition that can effectively use its experiences, the transcendental unity of apperception along with the cognitive faculties which Kant elaborates must be accounted for. This outlines the components necessary for the development of true artificial intelligence. The following sections detail further implications of this type of mental activity on creating intelligent machines.

3.3 The Categories of Judgement and Traceable Intelligence

Part of Kant's theory of cognition includes general categories under which synthetic judgements fall. Synthetic judgements combine information to create new knowledge, as opposed analytic judgements, which further elaborate on known information without creating new knowledge. Kant determined that judgements fall into twelve distinct categories, and these categories descend from categories within formal logic. Three of these categories are called judgements of relation and form the basis of Kant's analogies of experience. These categories of judgements are part of the a priori structure of consciousness. Dr. Christopher Tucker argues that if these categories of judgements are applied to machine consciousness, then it becomes "possible to introduce any set of controls on the program" (Tucker 3). Accomplishing this would require designing these a priori aspects of consciousness, such as the categories of judgements and spatio-temporal relationships, as inherent to the machine's cognitive structure. Current machine intelligence is unaware of why exactly it carries out tasks in the specific ways that it does; consequently, tracing artificial intelligence's behavior becomes impossible. That said, the ability to trace the behavior of artificial intelligence has serious ethical implications. If machines employ behavior without a traceable method of decision-making, artificial intelligence could make choices that are heinously unethical or could endanger human lives. Unresolved, such problems could also endanger the future of artificial intelligence. The ability to trace the behavior of artificial intelligence by providing controls and building structures into the very fabric of machine cognition would safeguard against those possibilities and lead to safer artificial intelligence that could benefit humans without harm.

3.4 Applying Kantian Causation to Machines

In section 3.3, I established that modern machines largely use a Humean type of causation. This means that they use correlation and probability to determine their actions. This type of causation is severely limited and inefficient in its function. An example given by Dr. Sloman to illustrate its limitation is seeing two interlocking gears. Humans who have interacted with similar situations before will instantly understand how the movement of one gear will influence the other. This is due to the mental structure and process of causation that Kant describes. I will refer to this manner of cognition as Kantian causation. Kantian causation, due to the way it applies concepts, such as gears and spatial relationship, to sensory data allows for understanding of causal structures even in novel situations. Even if the human subjects have never seen these specific gears before, synthesizing conceptual representations with sensory data allows subjects to gain knowledge. Since most machines employ Humean causation, which is equivalent to correlation, current machine intelligence could only deal with this problem of interlocking gears after much trial and error. This presents a further limitation whereby machines are unable to reason efficiently when dealing with unique or unfamiliar circumstances. Of course, Humean causation is necessary in the beginning of experience itself. This is because subjects need representations that will be reproduced by the imagination and fit within the semantic network of knowledge through the transcendental unity of apperception. However, after these incipient stages, Kantian causation is the only route forward. Using imagination to reproduce the intuitions and fit them within the a priori intuitions of consciousness leads to causal reasoning about structures. These are the necessary components for the design of an artificial intelligence that can learn, reason, and deal with new and increasingly abstract situations.

Portfolio Conclusion

This portfolio clarifies the limitations of current machine intelligence and outlines the elements of cognition necessary for reasoning machines. Understanding what the cognitive structure of a reasoning machine must look like is the first step toward creating it. The resolution of David Hume’s challenge to inductive reasoning provides insight on the path to a better model of reasoning than the correlation-based reasoning which limits current machines. Immanuel Kant’s transcendental idealism clarifies the cognitive faculties and structures which define experience and allow for causal reasoning and illuminates the fundamental principles which must be present in the structure of a reasoning and ethically traceable artificial intelligence.

Portfolio Bibliography

Athalye, Anish, et al. "Fooling Neural networks in the Physical World with 3D Adversarial Objects." *LabSix*, 31 Oct. 2017, http://www.labsix.org/physical-objects-that-fool-neural-nets/

This article examines Google's InceptionV3 image classifier and the ways in which it can be tricked. The author discusses the various ways in which machine intelligence can have its pattern recognition manipulated.

Casten, J.D. *Cybernetic Revelation: Deconstructing Artificial Intelligence*. Post Egoism Media, 2012. PDF.

Casten's book provides a thorough history of epistemological thought and draws parallels to artificial intelligence. The author provides excellent commentary on the translation of technical philosophical terms to more modern computer science terms.

Gardner, Sebastian. *Kant and the Critique of Pure Reason*. Routledge, 1999. Print.

Gardner's book provides an indispensable explanation of the complex theories of Immanuel Kant. The book comprehensively outlines the entirety of Kant's transcendental idealism and places it within its broader historical context.

Kant, Immanuel. *Critique of Pure Reason*. Edited by Marcus Weigelt. Translated by F. Max Müller, Penguin, 2007. Print.

This book is one of the most important works of Immanuel Kant. It provides the basis for transcendental idealism. Kant provides a comprehensive account of mental activity and knowledge construction.

Kant, Immanuel. *Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science: with Selections from the Critique of Pure Reason.* Edited by Gary C. Hatfield, Cambridge University Press, 2008. Print.

This book helps explain Immanuel Kant's transcendental idealism. This book provides important philosophical context and historical background for Kant's arguments.

Morrison, Joe. "Skepticism About Inductive Knowledge." *The Routledge Companion to Epistemology*, edited by Sven Bernecker and Duncan Pritchard, Routledge, 2011, pp. 445-453. Print.

Morrison's work summarizes the famous arguments of David Hume. This article also summarizes the major tenets of the inductive skepticism which David Hume sparked.

Schantz, Richard. "Skepticism and Anti-Realism." *The Routledge Companion to Epistemology*, edited by Sven Bernecker and Duncan Pritchard, Routledge, 2011, pp. 477-487. Print.

Schantz's article provides an excellent explanation of the tenets of epistemological realism and anti-realism. Schantz also outlines the necessary conditions for refuting epistemological skepticism.

Sloman, Aaron. "Understanding causation in robots, animals and children: Hume's way and Kant's way." September 2007. PDF.

Sloman's presentation explains how Kantian causation might be applied in the broader context of cognitive systems. This work also provides many useful examples of robots' relationships to causal interactions.

Tucker, Christopher A. "A proposal for ethically traceable artificial intelligence." PDF.

Tucker's paper analyzes how Kantian categories of knowledge could allow for an artificial intelligence with traceable behavior. This proposal has significant implications on the future of ethics in the context of artificial intelligence and machine learning.