A Pragmatist Perspective on Causation, Laws and Explanation

Richard Healey

[This paper was published in J. Voosholz, M. Gabriel (eds.), *Top-Down Causation and Emergence* (Springer Nature Switzerland AG 2021, pp. 253-268]

§1. Introduction

The subtitle of George Ellis's (2016) book *How Can Physics Underlie the Mind?* flags top-down causation as the key notion needed to answer the question posed by its title. This notion is in tension with the Laplacean vision of causation as the playing out of laws governing the global evolution of the world by determining the motion of its basic physical parts. Some physicists and philosophers still think of causation as ultimately physical, even if the fundamental laws are not deterministic. But causation plays a vital role through all of science and human life in ways that may be understood locally, and without reference to physical (or even any) laws.

I will offer a pragmatist understanding of causation, laws and explanation that traces the features of these notions to their functions in our practical as well as theoretical projects. Laws derive their importance from their epistemic and methodological functions, while the primary role of causal concepts is in guiding action. Contemporary interventionist accounts of causation and causal modeling appeal to and clarify this practical role while downplaying the causal significance of laws. They also explain how causation in one science or at one level of complexity may be either related to or independent of causation in other sciences or at other levels. In this way they can demystify the notion of top-down causation by showing how, and when, it is possible.

At the most general level, explanation leads to understanding: but this can take different forms. Consequently, explanation serves different functions in science and comes in many varieties. Some scientific explanations are causal, but others are not: many scientific explanations appeal to laws, but many do not. Unifying explanations serve the epistemic function of connecting otherwise separate items or branches of scientific knowledge. This promotes an economy of thought but can also strengthen the evidence supporting one or more hypotheses by adding links to newly relevant data. Causal explanations can be especially satisfying when they lead to the kind of practical understanding that permits control over the phenomena explained.

I will illustrate these general points by means of specific examples, including smoke detectors, measurements of the Hubble constant, and the game of Life. Finally I will comment on how far this pragmatist perspective helps one to see how physics can underlie the mind.

§2. Causation

I follow most philosophers in calling the relation a cause bears to its effect *causation*. The alternative term 'causality' has associations it is best to avoid. These emerge in a famous passage from Laplace (1820):

> "We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be

> uncertain and the future just like the past would be present before its eyes."

Known for his successful application of Newtonian mechanics to the solar system, Laplace here contemplates its application to all bodies of the universe. His second sentence has been understood to imply a thesis of universal determinism, to the effect that the entire history of the universe is determined by its present state. The first sentence may be read as a proposed analysis of the causal relation that would make it both global and deterministic, thereby supporting a "principle of causality" according to which every event has a preceding cause that determines its occurrence. Indeed, causality has sometimes been confused, if not *identified*, with determinism.

In fact, causal relations are typically neither global nor deterministic: the lightning caused the fire, the fire made the alarm go off, the sound of the alarm brought out the fire department, etc. To reconcile such examples with the Laplacean vision one would have to take the lightning, fire and sound each to be determined by a localized part of some global event and to appeal to exceptionless laws of time-evolution relating these global events. But with no access to such global events and no knowledge of the required laws we are nevertheless somehow in a position confidently to make such causal claims about the fire.

In fact basic physics gives us reason to believe there *are no* exceptionless laws of time-evolution relating the state of the universe when the fire breaks out to its state when the fire alarm sounds. A fire alarm is typically triggered by an ionization smoke-detector whose operation relies on the radioactive decay of a small sample of the Americium 241 isotope by emission of alpha-particles. These ionize enough of the air in a thin gap between metal plates to produce a small electric current in a circuit as the ions are collected on these electrodes. As the gap fills with smoke from a fire, some of the ions are removed by adhesion to smoke particles, the current decreases, and a switch in the circuit sets off the alarm. But radioactive decay is believed to be an indeterministic process: it is entirely possible for the current flow to be maintained because enough Americium nuclei randomly decay while the smoke particles are in the gap to keep the current flowing at the original rate, so the alarm fails to go off.

An ionization smoke detector works not because there are laws that relate the global state of the universe at different times, but because there is a robust regularity relating the current through its circuit to the presence of smoke in the gap in that circuit. Robustness here means that the regularity is preserved under a wide range of variations in other features of the detector and its environment, including the temperature of the air, the amount and composition of the smoke, and how long the detector has been in place. The regularity is not robust against other variations, including draining of its battery, placement under water or in a sealed container, and overenthusiastic frying.

This regularity is not accidental, but nor is it an instance of any law of nature, fundamental or otherwise (though its holding does depend on a stochastic physical law of radioactive decay). Underlying it is a functional relation between two magnitudes: an inverse relation between the amount of smoke in the gap between the electrodes and the current in the circuit. Contemporary interventionist approaches to causation represent these magnitudes as nodes in a Directed Acyclic causal Graph (DAG) (see Pearl (2009), Woodward (2003)). The direction of the graph corresponds to the direction of causation, so one can represent the fact that the smoke causes the sound of the alarm (not *vice versa*) by the orientation of the arrow linking their nodes on the graph. The orientation of a link in a DAG may be checked by controlled interventions on the nodes it connects. To test a fire alarm one can intervene by increasing the amount of smoke between its electrodes and look for a corresponding variation in the current through its circuit. Changing the current through the circuit by increasing the voltage across it will not change the amount of smoke in the gap in the circuit.

Such causal modeling techniques do not yield an analysis of causation insofar as the notion of an intervention is itself causal. Woodward (2003) gives a more precise account of an intervention in causal terms while insisting that this does not rely on any appeal to human agency. Price (2017), by contrast, locates the origins of the concept of an intervention within the perspective of a hypothetical agent able to contemplate the alternative possible consequences of his or her own decisions—a perspective that may itself depend on how the agent is physically situated. Either way, one can offer a non-reductive account of the causal relation between events *c* and *e* of types *C*, *E*, respectively in terms of counterfactual statements about the effects of an intervention on *c* that replaces it by an event *c** of type *C**, where this different type of event corresponds to the same magnitude taking on a different value.

Interventions may or may not be understood as natural occurrences independent of the perspective of any agent. But they are certainly just what one needs to attain goals by manipulating elements of the environment. In this way an interventionist approach to causation explains the practical function of causal concepts for agents like us—physically situated as we are in a small part of a much bigger and longer lasting universe, to which we have very limited direct epistemic access though our senses. This makes an interventionist approach to causation very attractive to a pragmatist whose first question when confronted by a problematic concept is to ask what function is served by possession and use of that concept. As we'll see in §6, it will also help to explain the nature and possibility of downward causation.

§3. Explanation

I begin with a rapid and selective review of recent history of philosophical thought about scientific explanation, emphasizing changing views on the roles of laws and causation.

Logical positivist philosophers and their logical empiricist heirs maintained that any fully satisfactory scientific explanation must include at least one scientific law (see Carnap (1966/1995, chapter 1), Hempel (1965)). Critics exhibited cases in which, they argued, laws sufficient to effectively guarantee the occurrence of the phenomenon in question nevertheless failed to explain it: what was missing in these cases, allegedly, was any claim about the *cause* of the phenomenon (Bromberger (1966), Salmon (1984)). Lewis (1986) argued that to explain the occurrence of an event was just to provide information about its causal history, and backed up this proposal with an influential analysis of causation (1973/1986) in terms of counterfactuals—statements of the form *If a had not happened then b would not have happened*. Following these developments, philosophers came to emphasize the importance of causation in understanding what it is to explain something in science without reference to laws. Some (such as Strevens (2008)) maintained that all scientific explanation is causal, while others (including Woodward (2003)) took their accounts of causal explanation as models to be extended to accounts of other kinds of scientific explanation.

But Maudlin (2007) argued, against this trend, that causation and counterfactuals are best understood by appeal to laws of evolution, whether fundamental (in physics) or derivative (as in special sciences). And more recent work by philosophers (Lange (2017), Reutlinger and Saatsi (2018) has explored a variety of different kinds of non-causal explanation in science.

Those with a jaundiced eye may see the fluctuating fortunes during this period of appeals to law and to causation as a glaring example of philosophy's failure to progress as a discipline. As a pragmatist I see here a case in which explanation has multiple roles in science that may be played by different actors. If this is right, we should not seek a unified theory of scientific explanation, but an account of a variety of explanatory tasks and a corresponding variety of kinds of scientific explanation best suited to accomplish these.

Some scientific explanations are prized because of their unifying power. In biology these include explanations that locate a species or other taxon on a single tree of terrestrial life, explanations in terms of natural selection, and explanations that appeal to a small stock of biochemical ingredients (such as nucleic acids and their bases) and processes (such as protein synthesis). In physics they include the explanation of heat as molecular motion and of light (as well as radio waves, X-rays, ...) as electromagnetic radiation within a certain range of wavelengths, as well as the common pattern of mechanical explanation that Newton used to unify celestial and celestial phenomena.

A unifying explanation need not be causal, and will not be if what it unifies are laws, theories or branches of science rather than natural phenomena. Unifying explanations serve methodological needs within science by helping to organize and simplify the structure of scientific knowledge.

Causal explanations are valued because they contribute to our practical knowledge. Some do so directly by showing how we may achieve practical goals of manipulation and control by appropriate interventions. But a causal explanation may yield understanding even of phenomena in which we cannot intervene. It can enable us to represent these phenomena within our perspective as agents situated in a world on which we can act, rather than as passive observers of events that play out in that world.

§4. Laws

While laws play several roles in science—especially physical science—it is fruitful to think of these as all related to a function pointed out long ago by Gilbert Ryle (1949, p. 121):

> "Law-statements are true or false but they do not state truths or falsehoods of the same type as those asserted by the statements of fact to which they apply or are supposed to apply. They have different jobs. At least part of the point of trying to establish laws is to find out how to infer from particular matters of fact to other particular matters of fact… . A law is used as, so to speak, an inference-ticket (a season ticket) which licenses its possessors to move from asserting factual statements to asserting other factual statements."

Ryle's distinction between two types of truth (statements of fact and the true law statements to which they apply) is murky. But his emphasis on the distinctive job of law-statements in licensing scientific inference is important and worth salvaging.

There is a lively debate in contemporary philosophy about an issue in the metaphysics of scientific laws: Do laws govern the construction of the physical world, or is the framework of scientific law a metaphysical superstructure supported by what actually happens in it? For a pragmatist this is the wrong question to ask: laws provide science with an epistemic infrastructure, not a metaphysical superstructure. Law statements do this by licensing reliable inferences that connect not only singular statements but also other law statements and, through them, different theories and branches of science. To play this role a law statement need be neither universal nor even true. But endorsement of a law-statement commits one to its general reliability in an acknowledged domain of applicability (its scope).

Astrophysics and cosmology provide some of the most impressive examples of the epistemic function of scientific laws in gaining new knowledge. Without their aid we would know very little about the solar system and next to nothing about what lies beyond it. Of course instruments such as telescopes have also played a vital role in acquiring knowledge of what lies beyond our physical reach. But scientists rely heavily on inferences warranted by scientific laws in their construction, calibration and deployment, and especially in the interpretation of the data they provide. Here are some examples.

We know a great deal about the atomic composition of stars and other celestial objects by telescopic observation of the light they emit or absorb. Such knowledge depends on inferences licensed by laws of spectroscopy. This is how helium was discovered on the sun 27 years before any was found here on the earth. In combination with laws of the Doppler shift (and subsequently general relativity) laws of spectroscopy license the inference to the conclusion that (on a sufficiently large scale) the universe is expanding in accordance with Hubble's law that recession velocity is proportional to distance. But justification for belief in Hubble's law also depends on knowledge of the distance to the receding body—knowledge that is come by only through inferences crucially involving additional scientific laws (such as the initially crude empirical generalization known as the Leavitt-Shapley law relating the period and absolute luminosity of Cepheid variable stars).

The exact value of the proportionality constant in Hubble's law is currently a matter of intense investigation because apparently reliable observational techniques that require inferences using different scientific laws lead to significantly different values. To extend the metaphor of epistemic infrastructure, it is as if engineers digging a tunnel under a mountain from different sides to connect a road or rail network had trouble precisely aligning their tunnels where they met under the mountain. The investigation will be successfully concluded when consensus is reached on which techniques and laws are to be relied on in determining the exact value of the Hubble constant at different epochs.

Scientific laws also license inferences that serve the function of standardizing and applying existing scientific knowledge. The 2019 redefinition of basic units in the metric system of units provides an interesting example. Four units were then redefined in terms of so-called constants of nature: the kilogram, kelvin, ampere and mole were defined by setting Planck's constant h, the speed of light in a vacuum c, the charge on the electron e, Boltzmann's constant k equal to specific numerical values when represented in these units. The assumption that each of these magnitudes is indeed constant is justified by inferences from basic physical laws into which they enter. Prior to the redefinition various techniques had been used to *measure* the value of each constant. This is no longer necessary since their values have been fixed by definition. Instead, these same techniques are used to measure the mass, temperature, current, or amount of a sample in the relevant units.

Each such measurement technique itself relies on a variety of scientific laws. For example, a watt or Kibble balance may be used to very accurately measure the mass of an object. A large number of physical effects are involved in the operation of the balance, some electromagnetic, others mechanical. Two effects are explicitly quantum mechanical, in the sense that laws of quantum theory are needed to use these effects to infer the values of relevant magnitudes: the Josephson effect and the quantum Hall effect together permit very accurate measurements of current and voltage, and hence electric power. Laws of classical mechanics, gravity, the Doppler effect, and elementary electric circuit theory are all also used to infer the mass of the measured object from the observed behavior of the balance as well as other devices in and outside the laboratory in which the balance is located.

These laws have very different pedigrees. Few, if any, are now regarded as fundamental: some long predate the 20th century relativity and quantum revolutions. Many are known not to hold in certain situations: they are neither true nor universal. The laws justifying an inference to the mass of the measured object may even be collectively inconsistent. But that inference is nevertheless licensed because each law is applied only where inferences based on it can be relied upon in reaching conclusions that are accurate within assessable error bounds.

I noted in the previous section that some scientific explanations appeal to laws while others appeal to causes. A good explanation of the operation of a Kibble balance would advert to many of the laws used to infer the mass of the measured object. But laws used to infer the value of the Hubble constant would not explain why the Hubble law holds: that calls for a causal explanation of why the universe is expanding in accordance with the law.

§5. Emergence and the Life World

Downward causation is sometimes thought of as a phenomenon in which emergent, higher-level properties causally effect lower-level properties out of which they emerge. But both emergence and downward causation are problematic concepts. In discussing them it will be helpful to have in mind a model that is simpler than the actual world but still rich enough to provide illustrations of potential analogs to emergence and downward causation. This is John Conway's game of Life.

Life is not so much a game as a kind of cellular automaton. Events in a Lifeworld occur within a two-dimensional grid of congruent square cells in which each cell touches eight neighboring cells. Events play out in discrete time steps. At each step a cell has one of two properties: it is in a state at which it is said to be either alive or dead. The state of a Lifeworld at step $t+1$ is uniquely determined by the distribution of these properties over its cells at step t. Here we assume a criterion for what counts as the same cell at different times. The determination relation is local: the state of each cell at $t+1$ is a function of its state and those of its eight nearest neighbors at t, as follows:

Any live cell with fewer than two live neighbors dies.

Any live cell with more than three live neighbors dies.

Any live cell with two or three live neighbors lives, unchanged, to the next generation.

Any dead cell with exactly three live neighbors will come to life.

One can think of the cells as micro-objects in a Lifeworld, at each time-step in a micro-state of being alive or being dead. The state of a cell at t constitutes a micro-event. One can also regard the local determination relation as specifying micro-laws of time evolution that hold for these micro-events.

This micro-structure generates a remarkably rich macro-structure of emergent objects, properties, laws and causation in some (though not all) Lifeworlds that begin in a particular global micro-state and then evolve in conformity to the above conditions. The possibilities are most easily explored by implementing the defining conditions as a computer program and watching how the chosen Lifeworld develops on the screen. Many such implementations are now readily available for download from the Web.[1] One basic macro-object is called a *glider*. At each time-step a glider is a shape made up of five contiguous live cells, each with no other live nearest neighbors. The glider changes at every time-step: the live cells cycle through three different shapes before the glider returns to its original shape. But the cells in that shape at $t+4$ are not the same as those that made it up at t: the original pattern has been shifted over diagonally, and is now formed by cells located elsewhere on the grid.

A glider constitutes a macro-object that retains its identity despite changes in its shape, location and component micro-parts. It may be taken as an example of an emergent object, and its shape and location as examples of emergent properties of that object. Its behavior is consistent with an emergent macro-law of motion specifying that it moves one cell horizontally and one square vertically every four time-steps. A single glider in an otherwise empty Lifeworld (all other cells are dead) continues to exist and to move in accordance with this law forever. But if there are other live cells its behavior may differ, as we shall see in the next section.

David Chalmers (2006) distinguished two notions of emergence as follows:

> "...a high-level phenomenon is strongly emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are not deducible even in principle from truths in the low-level domain."
>
> "...a high-level phenomenon is weakly emergent with respect to a low-level domain when

---

[1] Click on the following URLs to access programs implementing, respectively: glider, glidergun, glidergungliderdestruction

> the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are unexpected given the principles governing the low-level domain."

Life presents no examples of strong emergence because all high-level phenomena are determined by the starting micro-state in accordance with the local, algorithmic rules of the "game". But it will serve as an illuminating example of weak emergence, since a Lifeworld can display quite unexpected high-level behavior. The Lifeworld primer, for example, evolves so as to produce a stream of leftward-moving "space ships" with the spaces between them proportional to those between the prime numbers. Most strikingly, there is a Lifeworld that can be thought to realize a universal Turing machine and so implement any digital computer program. Such examples of weak emergence suggest this playful tweak to the conclusion to Charles Darwin's *Origin of Species*:

> There is grandeur in this view of Life,... from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.

Chalmers (2006) takes consciousness to present the only example of strong emergence because he believes that even if consciousness has physical correlates in our world these don't suffice to deduce their conscious correlates without additional psycho-physical laws that are not themselves consequences of laws like those physicists currently have or seek. His view remains controversial among philosophers of mind and the issue is unlikely to be resolved without taking a stand on the metaphysical status of laws―something that a pragmatist would do well to avoid.

Someone might appeal to quantum entanglement as an instance of strong emergence because the physical properties of compound systems are not deducible from those of their component subsystems. But this appeal fails if, as I believe (see my 2016) entanglement is not a physical relation.

§6. Downward causation

The metaphor of downward causation depends on a conception of the world as stratified into levels ordered from lower to higher, or even bottom to top. In one way of understanding it, the origin of such an order lies in a composition relation that relates objects to the wholes they compose. For example, quarks are often said to compose nucleons that form the nucleus of a carbon atom whose other parts are electrons, while the atom is itself part of a DNA molecule that is part of the nucleus of a neural cell in a human brain. Despite such examples it is not clear how well the composition relation and the levels conception itself stand up to critical scrutiny. This is not the place to provide it, so I'll assume they do and pass on to see how the metaphor of downward causation arises and why it has seemed problematic.

The levels conception suggests the possibility of reducing the behavior of an object to that of its component parts, where reduction is thought to permit or even require explanation. The possibility holds out the promise of another kind of order in which facts about objects at different levels are related by an explanatory relation. Steven Weinberg (1992, 2015), for one, sees science as succeeding in its attempt to fulfill this promise. In his (2001) essay, Weinberg himself proposes no well-developed theory of scientific explanation while encouraging philosophers in their attempts to clarify the notion. But he stresses the importance of explaining regularities or principles rather than individual events, and his remarks on deduction hark back to the logical empiricist theories of Hempel while he raises doubts about explanation in terms of causes.[2]

Strevens (2008), by contrast, gives causal explanation center stage even while agreeing with Weinberg that explanation is located out in the world, not in the communicative acts of scientists. For levels to be asymmetrically ordered by causal explanatory relations these must also be asymmetric: they must explain what goes on at higher levels in terms of underlying lower level causal mechanisms. A relation in which higher level objects or processes cause lower level events or phenomena might threaten the asymmetry inherent in this perspective. Indeed, Jaegwon Kim (e.g. 2005) has formulated so-called causal exclusion arguments intended to rule out the possibility of any downward causal relations. But a pragmatist who adopts the interventionist approach to causation discussed in §2 has a ready reply to such causal exclusion arguments (Woodward, 2015).

Consider a case in which the instantiation of a low-level property P_1 determines the instantiation of a distinct high-level property M_1 and P_1 is a sufficient cause of a different low-level property P_2 whose instantiation determines that of a distinct high-level property M_2. Here is a causal exclusion argument intended to show that M_1 cannot be the cause of P_2. By assumption, P_1 caused P_2. If M_1 also caused P_2 then P_2 was overdetermined, since P_1 was its sufficient cause and M_1, P_1 are not causally related. But this case is quite different from standard cases of overdetermination: all the "causal work" here is done by the lower-level cause P_1. M_1 is not an overdetermining cause but a mere epiphenomenon. Since this argument is quite general there can

---

[2] In a puzzling footnote Weinberg (2001) takes Kitcher to advocate a causal theory of scientific explanation. But Kitcher (1981) sees *unification* as the key to scientific explanation, apparently in agreement with Weinberg himself.

be no downward causation.

Roughly, on an interventionist approach M_1 is a cause of M_2 just in case appropriate interventions that change the property M_1 are accompanied by corresponding changes in M_2. This is a crude statement of the view expressed by Woodward (2003) who here supplies the missing details. When it is applied to the case of the previous paragraph this approach may or may not deliver the verdict that M_1 is a cause of M_2. As Woodward (2015) points out, it is consistent with the description in that paragraph to suppose that M_1 and M_2 are not even correlated properties. But one can point to cases meeting that description in which the approach will correctly count a high-level property as cause of another even though their instances are determined by instantiation of low-level properties P_1, P_2 respectively where P_1 is a sufficient cause of P_2. In such case M_1 will count as a cause of P_2 as well as M_2, so we have an unproblematic example of downward causation.

We can find many such cases in the game of Life. In §5 I claimed that some Lifeworlds contain a rich causal macro-structure. Here are some examples. A glider is a persisting macro-object that perdures by constituting a causal process in which each time-stage produces the subsequent stage of the glider. A glidergun is a macro-object that remains at the same macro-location while producing a sequence of gliders. In glidergungliderdestruction five gliders converge on a complex macro-structure and are successively destroyed while destroying that structure.

In the third example, consider an early period during which the first glider is destroyed as it encounters a stationary object composed of a block of four live cells that is also destroyed in the process. We may define two macro-properties of the Lifeworld: containing a glider moving south-east across the world, containing a small stationary block of live cells in the path of that glider. Each macro-property is determined at each time-step by the micro-properties “being alive”, “not being alive” of particular individual cells. It is intuitively clear that the first macro-property causes the second not to be possessed as time passes. This agrees with the verdict of the interventionist approach. An intervention that altered the Lifeworld just by giving the glider a north-east trajectory or turning it into a similar stationary block of live cells would ensure continued possession of the second property. That same intervention would also alter those micro-properties of the Lifeworld that determine possession or non-possession of this second macro-property. So this is a case of downward causation in a Lifeworld: a case in which a higher-level property causally influences lower-level properties.

The game of Life can provide a powerful intuition pump (to use Daniel Dennett’s term), but we do not live in a Lifeworld. Are there examples of downward causation in the real world? A contemporary LCD computer screen provides some. I am now composing this paper by pressing keys on a keyboard connected to a personal computer. As I press each key a black letter appears on the computer monitor. This is not a coincidence. Pressing the key marked D (as I just did) caused a black mark shaped like an upper case letter *d* to appear at a specific place on the screen which had previously appeared white. A sequence of such key-pressings caused the previous sentence to appear on the screen. Here we have familiar instances of macro-causation observable in the real world.

A black D-shape reliably appears on the screen when the key marked D is pressed thanks to the operation of unobservable micro-processes. Electrical currents produced by depressing the key are input to a word-processing program running on the computer. Its output causes electrical currents to pass through some tiny liquid crystals sandwiched between two crossed layers of polarizing material, all back-lit by white light. Three crystals compose each of the roughly 1million pixels arranged as a rectangular grid across the screen. Passage of electrical currents through all of the three crystals in a pixel allows horizontally polarized light to pass through them, only to be blocked by the second layer of vertically polarized material. So none of the light emerges from that pixel. That's true of all the pixels making up a D-shape at that place on the screen. So a black D-shape appears on the screen. No similar electrical currents pass through the crystals making up each surrounding pixel. Each of these crystals therefore rotates the plane of polarization of the light as it passes through so that light is now able to pass through the second layer of polarized material. That's why the surrounding screen looks white.

The continued appearance of the black D-shape on the screen is realized microscopically by the continual "refreshing" of the currents through the crystals making up the pixels in that area of the screen, 60 times each second. Every one of those micro-events was caused by my once pressing the key marked D on the keyboard. Together they determine the appearance of the black D-shape on the screen—the macro-event I caused by pressing the key on the keyboard. Here we have an actual example of downward causation. It would be easy to find many similar examples involving the use of contemporary digital computers.

If you ran the Life programs referred to earlier you have produced your own examples. Loading each program and pressing the Run key once causes the image on the screen to change in such a way as to represent the evolution of that particular Lifeworld. While a Lifeworld is itself a purely abstract structure, you can intervene in its physical *representation* by modifying which cells are alive at any particular stage in the evolution. In this way any instance of downward causation in a Lifeworld may be used to provide instances of downward causation in the real world.

Some have associated the downward-causal impact of high-level phenomena with the existence of autonomous high-level “configurational” laws (see McLaughlin (1993)). “Kicking in” only at a certain level of complexity, such macro-laws would supplement or even trump the operation of low-level laws. So a Laplacean intelligence knowing only initial conditions and micro-laws could not deduce the macro- or even micro-history of the world. The game of Life illustrates the fact that downward causation requires no such laws. Even though it manifests downward causation, its rules suffice completely to specify the evolution of a Lifeworld in terms of the behavior of individual cells. But there is still a sense in which robust macro-laws of Life are autonomous from these basic micro-laws.

On an interventionist approach to causation, there are many macro-properties in a Lifeworld that cause other macro-properties. The underlying correlation between their instances is robust against some interventions on the cause macro-property as well as against a range of variations in their environment. One who adopts the pragmatist view of laws promoted in §4 will be ready to call this correlation a macro-law. But, like all laws, it has a limited domain of applicability. The correlation will fail for interventions and/or environmental variations against which it is not robust. For example, adding a single live cell to a glider can make the glider self-destruct or evolve into a sequence of complex patterns that finally resolves into an alternating pair of nine objects instead of just gliding (try this using add to glider).

Such a macro-law is autonomous from micro-laws because there are cases in which it prescribes macro-behavior that cannot be realized if the micro-laws hold universally. In that case the macro-law fails to hold.   But, though subject to exceptions, the macro-law can remain of great practical benefit if (unlike in Life) you have no way of knowing the micro-state or micro-laws. In a sense, the micro-laws constrain the macro-law by determining its scope. If the micro-laws themselves are subject to exceptions in certain circumstances then there is even more room for the autonomy of macro-laws. In a pragmatist view the limited reliability of both laws and causal relations makes room for them to fulfil their theoretical and practical functions at any level and between any levels, whether upward, downward or horizontally.

§7. Program Explanation

Someone might claim that the real cause of a micro- or macro-property of the life-world is the program implementing that run. For example, the program primer outputs a line of left-moving "spaceships" representing the successive prime numbers 2,3,5,7,11,13,17,19,23, ... But as a set of instructions, a program is an abstract object like a number or string of symbols, and abstract objects have no causal powers. This remains true even when primer is implemented by running it on a physical representation of the required initial configuration of live and dead cells in a Lifeworld.

The idea that a program or something like it can be causally relevant to an effect without being causally efficacious in its production has been promoted by Pettit (1993, 2007) in his notion of program explanation. Perhaps a program such as primer can at least be causally relevant to the explanation of the production of a physical representation of a Lifeworld containing a line of left-moving "spaceships" representing the successive prime numbers 2,3,5,7,11,13,17,19,23,.... because it programs for this process? If so, one might maintain that the program itself is a downward cause of the micro-events that constitute the physical representation of the Lifeworld process involved.

I find this line of thought confused. §3's pragmatist view of explanation suggests a variety of different kinds of explanation of what happens when primer is run on a computer, and in particular of why the sequence of “spaceships” emerging from the left in a run of Life with initial data implementing that program is isomorphic to the sequence of prime numbers. Here are crude renderings of three kinds of explanation:

1. This follows from the four basic rules of Life and the initial Life-world configuration of live cells.
2. All spaceships corresponding to non-prime numbers have been eaten before they would have appeared in this sequence.
3. The program primer was designed to generate precisely this sequence of spaceships; acting on a physical implementation of the micro-structure of a Life-world it physically implements the four rules of Life when the program is run by using a computer mouse to press the “Run” button; and the “Run” button has just been pressed.

Explanations of these three kinds are not in tension with one another: none is “the real” explanation. The third sketches a causal explanation that begins with a macro-cause (masquerading as a teleological explanatory factor) but then proceeds to locate causes at the micro-causal level. The second gives a macro-causal explanation. If the term program explanation is to be used here it is most appropriately applied to the first kind of explanation. But that is not a causal explanation at all.

Here, as elsewhere, there may be program explanations, but the program itself is not a causally relevant factor in the explanation: at most it directs attention to the existence of causally relevant factors, though perhaps without specifying what they are (at least at the micro-level). The causal explanation itself does not mention any program, but only properties (in particular, values of magnitudes) and their causal relations, understood in terms of an interventionist model of causation. According to Pettit and Jackson (1990), a little reflection suggests that perhaps most of the explanations we are ever likely to offer will be program explanations, and (they presumed) we only reach potentially efficacious properties in physics at the level of unspecified micro-physical particles. In fact we can and do have plenty of causal explanations in daily life as well as throughout the sciences, few (if any) of which deal with specific micro-physical particles.

§8. The Mind

There is plenty of downward causation, but does it underlie the mind, and if so how? The issue of mental causation has motivated causal exclusion arguments intended to prove that mental properties have no causal powers of their own. By undermining such a causal exclusion argument the interventionist approach to causation reopens the question of downward mental causation—of physical events at the macro- or micro-level by the instantiation of mental properties. There seem to be clear cases of this. A few years ago a driver suffering an epileptic fit ran off the road and destroyed a row of saguaro cactus on the ranch where I live. Many other automobile accidents are caused by the mental states of drivers who are distracted by thoughts about other things. Discussion of mental causation tends to focus instead on cases of intentional action, where a physical event would not have occurred if an agent had not intended its occurrence. But it is not so clear that the relation is causal between an intention and its successful execution, or the consequences of its execution. Even the view that an action is caused by a collection of beliefs and desires has its critics.

In his book Ellis (2016) takes plans and theories as causes of physical effects, mediated by the minds of agents who execute or use them.

> "These plans and theories are all consequences of our thought processes. They are the result of top-down effects from abstract ideas to neural excitations and into the world, down to the level of atoms. We are surrounded by proof of the efficacy of mental thought. This effectiveness is based specifically on models and theories." (2016, p. 358)

This quote leaves the causal direction ambiguous. If a plan or theory is a *causal* consequence of our thought processes then can it also be a *causal* result of the plan or theory as an abstract idea? Ellis goes on to reject the (strange) idea that a theory is a single person's brain state, identifying it instead with an equivalence class of mathematical representations (which may themselves be represented physically—by writing them down, for example). As such, a theory is an abstract object which, as noted earlier, can be neither cause nor effect of any mental state(s) since it cannot be intervened upon.

One can certainly appeal to plans, models and theories in an explanation of the existence and features of a physical object like an Airbus (Ellis's example). But the explanation will not be causal: these *abstracta* do not cause it to exist, though physical representations of them clearly do play a causal role in the construction of the relevant physical object. They are able to play this role because of their effect on the mental states of the agents who construct them. Those mental states are distal downward causes of the presence and interconnections of the various atomic components of each micro-chip in the Airbus's control systems. The detailed conceptual model of an Airbus Ellis represents in his Fig. 7.14 might naturally be thought to supply the program for the construction of the aircraft in something like a program explanation.

§9. Conclusion

In order to understand the world, we need to deploy multiple explanatory strategies, each suited to the reliable but not exceptionless regularities that emerge at different scales or in different contexts. Not all of these yield causal explanations. A good explanation will yield understanding by unifying apparently unrelated phenomena. Sometimes we may be lucky enough to unify explanatory strategies at different levels by showing how phenomena at one level are reducible to, determined by, or grounded in phenomena at another level. In any case, we continue to seek explanations wherever we can find them. In theoretical science we treasure unifying explanations, whether or not these are causal. Practical life prioritizes causal explanations.

For a naturalist, there are two senses in which physics underlies not only the mind but every phenomenon studied by science. The first sense is just that if there were no physical things there would be no minds: naturalism precludes disembodied spirits. The second sense has to do with the scope of physics. Physics claims the entire natural world as its domain of applicability, unlike other sciences which focus on phenomena that manifest themselves only in special circumstances—chemistry only where in space-time there are atoms and molecules, biology only where there is life, neurophysiology only where there organisms with brains or at least central nervous systems, psychology only where there are agents with mental states. But each science retains a certain autonomy for its laws and explanations (whether or not these are causal): different specialized concepts prove useful in understanding and intervening in the physical world in different domains. Phenomena (weakly) emerge that are best described and understood using these specialized concepts. This is true even within physics. The Standard Model of elementary particle physics does not help us understand high temperature super-conductivity.

References

Bromberger, S. (1966) ‘Why-Questions’, in R. G. Colodny, ed., *Mind and Cosmos: Essays in Contemporary Science and Philosophy, volume 3*. Pittsburgh: University of Pittsburgh Press.

Carnap, R. (1966) *Philosophical Foundations of Physics*. New York: Basic Books. Reprinted (1995) as *An Introduction to the Philosophy of Science*. Minneola, New York: Dover.

Chalmers, D. (2006) ‘Strong and weak emergence’, in P. Clayton and P. Davies, eds., *The Re-emergence of Emergence*., 244-56. Oxford: Oxford University Press.

Ellis, G.F.R. (2016) *How Can Physics Underlie the Mind*? Berlin: Springer.

Healey, R.A. (2016) ‘A pragmatist view of the metaphysics of entanglement’, *Synthese*. Published first online 21 September 2016. DOI 10.1007/s11229-016-1204-z

Hempel, C. G. (1965) *Aspects of Scientific Explanation*. New York, NY: Free Press.

Kim, J. (2005) *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Kitcher, P. (1981) ‘Explanatory unification’, *Philosophy of Science 48* (4), 507-531.

Lange, M. (2017) *Because Without Cause*. Oxford: Oxford University Press.

Laplace, P. (1820). *Philosophical Essay on Probabilities*. Republished 1951, New York: Dover.

Lewis, D. K. (1973) ‘Causation’, *Journal of Philosophy 70* (1973), 556-7. Reprinted with postscripts in Lewis (1986).

Lewis (1986) D. K. *Philosophical Papers, Volume 2*. New York: Oxford University Press.

Maudlin, T. (2007) *The Metaphysics Within Physics*. New York: Oxford University Press.

McLaughlin, B. (1992) ‘The rise and fall of British emergentism’, in Beckermann, Flohr, and Kim, eds. *Emergence or Reduction?* Berlin: Walter de Gruyter.

Pearl, J. (2009) *Causality*. 2nd edition. Cambridge: Cambridge University Press.

Pettit, P. (1993) *The Common Mind*. New York: Oxford University Press.

Pettit, P. (2007) ‘Joining the dots’, in Brennan, Goodin and Smith, eds., *Common Minds*. Oxford: Oxford University Press.

Pettit, P. and Jackson, F. (1990) ‘Program explanation: a general perspective’, *Analysis 50* (2), 107-117.

Price, H. (2017) ‘Causation, intervention and agency: Woodward on Menzies and Price’. In Beebee, Hitchcock, and Price, eds., *Making a Difference*. Oxford: Oxford University Press, 73-98.

Reutlinger, A. and Saatsi, J. (2018) *Explanation Beyond Causation*. Oxford: Oxford University Press.

Ryle, G. (1949) *The Concept of Mind*. London: Hutchinson’s University Library.

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Strevens, M. (2008) *Depth*. Cambridge, Mass.: Harvard University Press.

Weinberg, S. (1992) *Dreams of a Final Theory*. New York: Pantheon.

Weinberg, S. (2001) ‘Can science explain everything? Anything?’, *The New York Review of Books*

Weinberg, S. (2015) *To Explain the World*. New York: Harper Collins.

Woodward, J. (2003) *Making Things Happen*. Oxford: Oxford University Press.

Woodward, J. (2015) ‘Interventionism and causal exclusion’, *Philosophy and Phenomenological Research 91*, 303-47.