Causation, Causal Perception, and Conservation Laws

Charles Twardy
History & Philosophy of Science
and
Cognitive Science
Indiana University
ctwardy@indiana.edu

Contents

Chapter 1

Introduction

> Thus the observation of human blindness and weakness is the result of all philosophy, and meets us, at every turn, in spite of our endeavors to elude or avoid it. Hume, *Inquiry*

This project is an experimental epistemology of causation in the spirit of Hume. My central questions and motivations are essentially those of Hume. So are my methods. Hume got philosophy into a bind over causation by bringing in empirical considerations. Two centuries later, empirical considerations can help extract philosophy from those very problems.

This thesis advances three claims:

- that causation does not involve necessary connection
- that causal connections can sometimes be perceived, and
- that a proper study of causation requires attention to the empirical details of perception.

I take as given that the goal of science is to provide understanding, and that it does this by giving explanations, some of which must be causal explanations, because causation forms the basis for our understanding of the world. The reason the Bell phenomenon in quantum mechanics is so disturbing is that there is no way to give a satisfying causal account, at least not without violating relativity theory. If causation plays a central role in science, then indeed it deserves its central place in philosophy and especially the philosophy of science. Because understanding is a human-centered concept, causation touches on psychology as well. However, there are more fundamental reasons for the interaction.

Causation is central to our understanding of the universe, and the identification of physical causes plays a key role in separating science from other branches of inquiry. But what is causation? We know some of what it is not. Most famously, causation is not mere correlation. Hume's insight was to explore our attribution of causality: when do we see some event in terms of cause and effect? Hume answered that often-repeated conjunctions of two events, one prior to the other, led us to expect the second when we saw the first. If there were few exceptions, we would feel that the connection between the two was necessary, or nearly so.

Hume's account lets in too much. There are universal generalizations which are not causal relations, from "day follows night" to "there are no diamonds with mass greater than 500 kg". The account can be patched, but even Mackie's INUS theory, widely regarded as the pinnacle of Hume-like theories of causality, suffers counterexamples stemming from Hume's basic insistence on necessary connection. Indeed, even an appeal to laws is insufficient, since there are plenty of purely functional laws such as the ideal gas law and even dynamic equations. Any two events A and B with a common cause C may likewise be related in a lawful way, even though there is no direct causal connection between A and B.

So we need to move beyond Hume's impasse. What *subset* of universal generalizations counts as causal? Like Hume, I contend that a satisfactory answer must meet and explain human perceptions of causality, since the category of causality is partly rooted in the way humans understand the world. In addition, following recent theories in philosophy of science, I expect that any satisfactory answer will correspond to something specifiable and perceivable in the outside world. There are objective relations in the world, but we get at the world through perception, action, and cognition. The endeavor to understand causation thus involves psychology in a central way.

Since in particular I am trying to answer the problem that Hume set, in a tradition whose essential structure was set up by Hume, I have to begin with an analysis of Hume's arguments. As it turns out, Hume's own account grows out of his understanding of the science of his day, which is to say Newtonian mechanics, and his Hume's aspirations to found the "moral sciences" on the same experimental foundation. My analysis of Hume's argument and the essential way it involves physics and psychology will serve well to understand the debate in modern philosophy, which is still conditioned by Hume's formulation, or various interpretations thereof. Indeed, it is haunted by Hume's ghost.

Psychology also traces its concern with causality back to Hume, but is usually more dismissive of Hume's arguments and motivations. Unfortunately, the twentieth century has not seen a great collaboration between psychology and philosophy. I say unfortunately, because progress in twentieth-century psychology and philosophy has opened up a real possibility for reformulating Hume's problem in such a way that a unified answer can be found to both his epistemological and ontological concerns.

What counts as causal is partly conditioned by the way we perceive the world. There are true causal relations which humans have a hard time seeing (moon and tides, food and health), and there are relations we tend to see as causal which are not (certain superstitions form one class). In some cases repeated exposure fails to make the relationship salient; in others, overly hasty generalization leads to a mistaken attribution of causality. Nevertheless, our refined scientific ideas of causality have grown out of ordinary perceptual causality, and though perception can be tricked, the fact that it has enabled us to survive for so long suggests that for the most part it has latched onto relevant causal regularities of the world. What regularities might these be? The only way we can answer that question is through empirical work. In addition to mounting a sustained defense of the importance of empirical work in philosophy, I present two new experiments designed to explore the link between the psychology of causal perception and the conserved-quantity theory of causality. I want to emphasize that by getting the psychology right, we can give ourselves strong grounds for dissolving some of Hume's problems.

Chapter 2

Hume and Hume's Background

David Hume had aspirations to create a systematic philosophy of the moral sciences, explicitly modeled on the natural philosophy of Isaac Newton. Although he later abandoned his attempt at a single system, he continued to press its various goals separately [87], embracing the empirical approach championed by Newton and Locke, but undermining their basis for drawing theological conclusions from empirical science.

Hume attempted to remove only that part of Newtonian science which allowed it to spread into theology, while preserving its ability to discover the laws of the universe. His primary tool was the analysis of the limits of human reason: Hume applied the methods of empirical science to scientific inquiry itself. His endeavor was partly psychological and partly philosophical, and both disciplines claim Hume, although they do not necessarily agree on what he said.

If we are to understand Hume's philosophy and the debates it inspired, we must view it in the context of the dawn of Newtonianism. Therefore, in order more closely to examine Hume's views on causation and the philosophy of science, this chapter pays special attention to the Newtonian features in Hume's philosophy, and the Newtonian atmosphere in which he wrote, and to which he responded. We will begin with Newton's own views on causation, the views of Newtonians such as Locke and Maclaurin, and those of anti-Newtonians such as Berkeley.

2.1 Newton (1642-1727)

Newton's *Principia* undermined the previous Cartesian mechanical philosophy by giving a mathematical account of force and motion *without* respect to the nature of the force. In fact, it seemed impossible to reconcile Newton's hypothesis of universal gravity with the mechanical world picture. Newton himself proved that Descartes' vortices were incompatible with both Kepler's laws and the inverse-square law. Worse still, for mechanists, universal gravity seemed to require instantaneous action at a distance, a concept Newton himself found abhorrent, if unavoidable.

Newton defended his use of unexplained forces by limiting the scope of natural philosophy itself to the mathematical description of motions and forces without regard to their cause. Despite that official stance, Newton was deeply interested in causes—which ultimately he identified with God—but exceptionally cautious about publishing views which he was unable to prove, and which were probably heretical. It is easier to understand Newton's ideas of causation if we divide them into two levels. On the first level, Newton gave the causes of phenomena. Newton demonstrated that white light is composed of many colors of light, and that prismatic refraction is caused by the differential refraction of light of different colors. On the second level, he did not have an explanation for why light of different colors refracted differently, and refrained from the mechanical hypotheses offered by so many of his critics. On the first level, Newton explained planetary orbits as motion under the influence of a universal inverse-square attracting force. On the second level, his analysis did not depend on knowing what the cause of gravity was; at different times he defended innate attraction at a distance, continual impacts from without, or the direct action of God. The explanations of the first level were what Newton referred to as true causes; those of the second level he often condemned as hypotheses which went beyond the available evidence.

Hume was very much taken with the idea of limits to human knowledge, and credited Newton's success to his recognition of such limits, and his unwillingness—at least officially—to extend inquiry beyond those proper bounds. Hume's skepticism was in part a generalization of Newton's caution, so we explore Newton to aid our understanding of Hume.

2.1.1 Newtonian Method

In his two-volume historical survey, *Causality and Scientific Explanation* [112, 113], William Wallace argued that causation has since Aristotle been held to be a central concern for science. In part, science has distinguished itself as a field by its search for causes. Wallace opposed the popular view that science is only about description, not explanation, and that Galileo and Newton succeeded because they ignored causes and explanations. Newton set himself apart from the past by insisting on the primacy of observation and experiment over hypotheses. Yet in much of his writing there is evidence that he wanted from science more than description. The Queries in his *Opticks* indicated that he was interested in the nature of light and gravity, among other things. Westfall [116] saw Newton's massive work on alchemy as evidence of his search for the causes of attractions and repulsions. Newton was interested in the nature of the causes operating in the world, but transformed science by postponing those questions in order to investigate the observable properties of the as-yet-unknown causes.

This move stood the reigning Cartesian philosophy on its head. Descartes held that the world was full of passive matter devoid of all qualities save extension and motion. In order to account for the rich world of experience, he spun out ingenious explanations of light, color, magnetism, and gravity using only the tools of his austere ontology. Light was the way our eyes interpreted the pressure of subtle particles, and magnetism was caused by the motion of screw-shaped particles through screw-shaped holes in iron. Descartes tested neither his hypotheses nor the real existence of phenomena he explained. But for any known phenomenon, Descartes could find a mechanical explanation.

That was exactly the kind of speculation Newton famously rejected in his, "I feign no hypotheses." Newton formed plenty of hypotheses, but the kind he detested were the speculative theories about matter so common among mechanical philosophers. Descartes and other mechanists had shown that it was possible to explain a rich and intricate world with the barest of fundamental notions, but they did not test that their explanations were correct. Newton did. In the eighteenth century and nineteenth centuries, philosophers including

Hume imitated Newton's methods in hopes of similar success.

Potential followers relied heavily on Newton's four "Rules of Reasoning in Philosophy" (*Principia* Book III), formulated to limit mechanical speculations and to encourage the pursuit of not just any agreeable causes, but the true causes. For reference, here is a list of the four rules as they stood in the third (1726) edition of the *Principia*. [108, p.5]

Rule I *We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.*

To this purpose the philosophers say that Nature does nothing in vain, and more is in vain when less will serve; for Nature is pleased with simplicity, and affects not the pomp of superfluous causes.

Rule II *Therefore to the same natural effects we must, as far as possible, assign the same causes.*

As to respiration in a man and in a beast; the descent of stones in Europe and in America; the light of our culinary fire and of the sun; the reflection of light in the earth, and in the planets.

Rule III *The qualities of bodies, which admit neither intensification nor remission of degrees, and which are found to belong to all bodies within the reach of our experiments, are to be esteemed the universal qualities of all bodies whatsoever.*

... [long explanation omitted]...

Rule IV *In experimental philosophy we are to look, upon propositions inferred by general induction from phenomena as accurately or very nearly true, notwithstanding any contrary hypotheses that may be imagined, till such time as other phenomena occur, by which they may either be made more accurate, or liable to exceptions.*

This rule we must follow, that the argument of induction may not be evaded by hypotheses.

Rule IV, which will figure prominently in our discussion of Hume's program, was added only in the third edition of the *Principia*, as a less subtle way to bar the speculative mechanical explanations Newton found so pernicious. [1] Newton's own activity did not always conform with his formal rules or statements of method, in the way that discovery scarcely conforms to the rules of justification, but he tried carefully to adhere to them in his published work.

However, ultimately Newton thought that God was the underlying cause, and even thought God's action could be inferred from general induction. Universal gravity, for instance, was unlike any other principle of nature. It extended forever and affected bodies according to their total quantity of matter, rather than just their surface. Such universality, and perhaps his inability to make any other explanation fit, led Newton to suspect that God himself was the motive power behind gravity.

While Hume adopted a general Newtonian methodology, he proved to be much stricter about the application of the fourth rule of reasoning. Hume saw strong parallels between the Cartesian mechanical hypotheses rejected by Newtonians and the Newtonian God hypothesis advanced by them. Rather than

[1] Recall when reading the rules that under the Cartesian philosophy, any hypothesis which fit the data was justified.

just argue that they were wrong, Hume sought to show that Newtonian natural theology violated fundamental tenets of Newtonian natural philosophy. We cannot evaluate Hume's claim without a deeper understanding of the Newtonian method, including the roles of hypotheses, mathematics, and causes.

Hypotheses

> And I shall not mingle conjectures with certainties. Newton

Newton's famous methodological dictum, *"Hypotheses non fingo,"* "I feign [or frame] no hypotheses" conjures up instrumentalist notions of a scientist describing "just the facts," without any metaphysical baggage. Newton's ultimate goal was not so austere, although he accepted relative sparseness as an intermediate step. Newton realized that the Cartesian method of hypotheses guaranteed only consistency, and not truth. Indeed, a clever philosopher could always invent some mechanism to explain a known effect, or another if that one proved unfit. Newton demanded more of explanations. He preferred first to establish the character of the phenomenon under investigation, as precisely and fully as possible. It was his hope that he could go that far without committing himself to a particular hypothesis. Afterwards, there would be room for hypotheses, *provided* they were of the sort that could be "directly and conclusively verified." [15, p.347]

Newton appears to have argued at one point the *System of the World*, a first draft of book III of the *Principia*, that his gravitational law was only a mathematical device, not a physical hypothesis:

> We said, *in a mathematical way*, to avoid all questions about the nature or quality of this force, which we would not be understood to determine by any hypothesis; [85, p.550]

However, I.B. Cohen has pointed out that existing Latin copies of Newton's text do not support the standard (Motte & Cajori) translation with its strong philosophical emphasis on mathematical *versus* physical investigation. Instead, Cohen rendered the passage as something closer to, "and to investigate mathematically its effects in moving bodies; further, in order not to determine its type hypothetically, ..." [19, pp.342-3]

Either way, there are two issues here, corresponding to the two levels of causation mentioned earlier. First, did Newton regard his forces as real? Yes. In the General Scholium of the third edition, right after "I frame no hypotheses," Newton stated that gravity was a real force:

> And to us it is enough that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of our sea. [84, p.547]

Chaudhury seems to have moved from a defense of the reality of gravity for Newton to the statement that Newton did entertain hypotheses. [15] In modern terms, that is correct. Even in his own time, Newton's opponents did not see the distinction. However, in his own view, Newton did not adopt the kinds of hypotheses he so clearly rejected.

Newton regarded universal gravity as a firmly established true cause, not a hypothesis. Newton's numerous "anti-hypotheses" passages were condemnations of speculative ultimate explanations, but not true causes. So Chaudhury was correct to say that Newton did not regard universal gravity as a mere mathematical fiction ("gravity does really exist"), but wrong to say on that basis that Newton entertained hypotheses of the sort he abjured.

Admittedly, the line between mathematical fiction and real existence is fuzzy. Knowing that ultimately, Newton underwrote universal gravity through the action of God, we can make sense of his assertion of its reality. It is harder, however, to understand the distinction between a universal description (all bodies accelerate toward each other as if attracted by an inverse-square force) and the further assertion that there is such a force, about which we can say nothing.

There are three inferences here for Hume to worry about.

- First, the inference from observed motions of a few bodies to *universal* gravitation.
- Second, the assertion that not only is this mutual acceleration universal, but that it is caused by some force.
- Third, the claim that this cosmology provides evidence for the action of God.

Hume's general skepticism held that none of the inferences were justified. However, according to De Pierris [22], within the realm of natural philosophy, he allowed the first, but thought that the second and third went beyond the bounds of reason. In this he differed from Newton, although it is not clear that he thought so.

Newton thought the hypothetical method was flawed, but that he had replaced it with the true method of analysis and synthesis. This latter method was designed to yield not just compatible explanations, but the true causes, in the same way that it guided the mathematician to true principles of algebra or geometry. Newton seems to have thought that if he followed this method, he could guarantee that any solution he arrived at was the true one. In this way he could maintain his rejection of mechanical speculation while allowing himself room to advance claims which others mistook for mere hypotheses.

Analysis and Synthesis

> By this way of analysis we may proceed from compounds to ingredients and from motions to the forces producing them, and in general from effects to their causes and from particular causes to more general ones, till the argument end in the most general. This is the method of analysis; and the synthesis consists in assuming the causes discovered and established as principles, and by them explaining the phenomena proceeding from them and proving the explanations. (Opticks, in [108, p.178])

This passage from the *Opticks* should make it clear that Newton was after causes, and that he thought he had a method for reaching them. Unlike Descartes, Newton allowed that forces were real, and cause enough for natural philosophy. According to him, the method of analysis and synthesis was particularly useful for finding these sorts of measurable but perhaps ultimately unfathomable causes. The reason was that its mathematical structure reduced the problem at hand to one of solving for the term representing the unknown force. The precision required to fit the numerical values guaranteed a level of agreement not possible with Descartes' qualitative hypotheses. However, an appreciation of the difference between the methods required an appreciation of the nature of mixed mathematical inquiry, something which appears not to have been one of Hume's strong points.

Analysis and synthesis itself was not a new method. Martha Fehér compared Newtonian and Aristotelian methods of analysis and synthesis, and their implications on the status of causal explanations. [32] While Newton desired to explain a change of state, Aristotle had looked for categorical belonging. Consequently, the link between the thing to be explained and the explanation was for Aristotle conceptual or logical, while for Newton it was mathematical. The move from categorical or conceptual to mathematical links made for very different kinds of explanations. Aristotelian explanations explained the lesser known by the better. Newtonian causes, however, were no longer observable members of the empirical domain, but unseen and unknown forces, unable in principle to explain anything on Aristotelian grounds. It was partly because of this shift in the nature of explanation that Newton's program was so consistently misunderstood, and Newton himself accused of traffic in occult causes. Without acknowledging Newton's claims for the method of analysis and synthesis, Hume would nevertheless provide a defense for the Newtonian method by showing that even the mechanists' apparently obvious causes were as unknown and unknowable as gravity.

What led Newton to this new conception of causation and explanation? Fehér tied Newton's introduction of abstract causes to his use of algebra, or 'the Analytic Art'. Indeed, the whole terminology of analysis and synthesis was mathematical. For Aristotle as for Newton, empirical analysis consisted in proceeding from known effects to unknown causes, and synthesis was the reverse. However, for Greek mathematicians from Eudoxus on, the method of analysis meant proceeding from unknown (but assumed) premises towards known theorems. [32, p.71] How did mathematics fit into Newton's conception of method?

Isaac Barrow, Newton's predecessor in the Lucasian chair, thought that only mathematical demonstrations were truly causal, because, "it is only formal causality which represents a truly necessary connection between the cause and the effect." [32, p.78] Since Newton was describing formal causes between measured magnitudes, he,

> could have felt himself justified in regarding his whole system as necessarily true and *unique*. .... He did not regard these laws as having merely a deductive *systematizing power* with respect to lower level (phenomenological) laws but as having a *causal explanatory* role as well." [32, p.79]

Indeed, Newton's exasperated replies to critics of his first paper on optics support that point of view.

Newton believed in deduction from the phenomena, according to the method of analysis-synthesis. Mathematical relations demonstrated to follow one from the other were, according to Barrow, causally connected. Newton seems to have subscribed to a similar view, since he he held that he had demonstrated his true causes with the force of mathematical proof.

2.1.2 Optical phenomena

In his 1671/72 letter to the Secretary of the Royal Society announcing his optical experiments, Newton was very clear about the status of his theory. Wallace noted the following section of that letter which did not get published in the *Philosophical Transactions*:

> A naturalist would scarce expect to see the science of those [colors] become mathematical, and yet I dare affirm that there is as much certainty in it as in any other part of optics. For what I shall tell concerning them is not an hypothesis but most rigid consequents, not conjectured by barely inferring 'tis thus because not otherwise, or because it satisfies all phenomena (the philosophers' universal topic), but evinced by the mediation of experiments concluding directly and without any suspicion of doubt. [111, pp96-97] in [112, p.196]

In the published part of the letter, Newton described in detail the shape of the spectrum cast by a prism, and the series of experiments by which he had disproved the accepted Cartesian explanations of refraction through a prism. Finally, he described his crucial experiment:

> And I saw by the variation of those places that the light, tending to that end of the image toward which the refraction of the first prism was made, did in the second prism suffer a refraction considerably greater than the light tending to the other end. And so the true cause of the length of that image was detected to be no other than that *light* consists of *rays differently refrangible*, which, without any respect to a difference in their incidence were, according to their degrees of refrangibility, transmitted toward divers parts of the wall. (*Phil.Trans. v.80* in [108, pp.71-72])

After a discussion of the implications of this discovery for telescope manufacture, Newton argued that further experiments still further demonstrated the heterogeneity of white light. Where Descartes had held that color is a property added to light, Newton showed that light was a mixture of light of different colors.

Merely disagreeing with Descartes did not make Newton an anti-mechanist. What set his method apart was that he stopped at that point and insisted that his formulation was as basic an explanation as it was possible at that time to give. He did not offer a mechanistic explanation for *why* colors are refracted differently.

> But to determine more absolutely what light is, after what manner refracted, and by what modes or actions it produces in our minds the phantasms of colors, is not so easy. And I shall not mingle conjectures with certainties. [108, p.78]

Although Newton did not produce a mechanical explanation of prismatic refraction, he thought he had discovered the "true cause" of the phenomenon, that white light is a mixture of different colors, each color associated with its own degree of refraction.

His critics pressed for a mechanical explanation, which Newton repeatedly stated had not been his aim. Indeed, Newton tried very hard to stay one level removed from the hypotheses of ultimate constitution, and to make his theories neutral with respect to such,

> for the best and safest method of philosophizing seems to be, first to inquire diligently into the properties of things, and establishing those properties by experiments, and then to proceed more slowly to hypotheses for the explanation of them. (Phil.Trans.v.84, p.4087, in [112, p.201])

Note that Newton did not reject the place of hypotheses in natural philosophy altogether, but only advocated that they be investigated slowly, and only after the properties of things had been established.

In the *Opticks*, Newton could claim that he had demonstrated many of these properties by experiment. However, in the *Principia*, he had to introduce ideas which were far more problematic, yet which he defended just as strongly.

2.1.3 Universal gravity

I. B. Cohen has argued that it is the third law of motion which led Newton to the idea of universal gravity.[2] The third law requires mutual attraction, hence neither planetary body was a fixed focus. Mutual attraction is significant effect in the sun-Jupiter system and the earth-moon system. Further, if each planet has this dual relation to the sun, the idea of planetary interactions suggests itself. "Newton concluded that one and the same force acts on each of the planets, because the planets are all of the same kind," in accordance with his first two rules of reasoning. [20, pp.35-6][3]

Cohen described the steps in the *Principia* by which Newton proceeded from the familiar towards universal gravity. I see six steps in Cohen's analysis:

1. We are all familiar with the phenomenon of weight, though we do not know the cause. Call the cause gravity.
2. A calculation shows that this terrestrial gravity may extend as far as the moon. (Prop.4, Bk.3)
3. So it can go from the earth to the sun as well.
4. And the sun to the planets.
5. And the planets to their satellites.
6. Gravity is universal.

After asserting universal gravity, Newton presented the evidence. Universal gravity explained the tides and the precession of the earth's axis, the common acceleration of all objects in the absence of air resistance, and the orbits of the planets.[20, p.38] All of this fit the strictures of *mathematical* principles of natural philosophy.

Newton wanted also to find some sort of physical cause, but nothing he thought of worked.

---

[2] As Cohen noted, this is more evident in Newton's *System of the World*. Consider:

> [20.] *The agreement of those analogies.*
>
> Since the action of the centripetal force upon the bodies attracted is, at equal distances, proportional to the quantities of matter in those bodies, reason requires that it should be also proportional to the quantity of matter in the body attracting.
>
> For all action is mutual, and (pp. 13, 26, by the third Law of Motion) makes the bodies approach one another, and therefore must be the same in both bodies. It is true that we may consider one body as attracting, another as attracted; but this distinction is more mathematical than natural. The attraction resides really in each body towards the other, and is therefore of the same kind in both. [84, p.568]
>
> [21.] *Their coincidence*
>
> And hence it is that the attractive force is found in both. The sun attracts Jupiter and the other planets; Jupiter attracts its satellites; and, for the same reason, the satellites act as well one upon another as upon Jupiter, and all the planets mutually one upon another.

[3] Newton is explicit about this in paragraph 22. See [84, p.569].

> Newton again and again sought for some explanation of how universal gravity might act. That is, he attempted to reduce universal gravity to the action of something else: a shower of aether particles, electrical effluvia, or 'spirits emitted,' variations in the density of an all-pervading aether. All of these attempted 'explanations' or reductions of universal gravity to some accepted kind of mechanism failed. None could fulfill two major requirements: that the resultant force vary inversely as the square of the distance and that this force act mutually on every pair of bodies so as to attempt to bring them together. [20]

In the end, he explained motion via attractions and repulsions, and such motion was amenable to mathematical analysis. However, despite much searching, he did not know what was responsible for those forces. He had ruled out the æther and any sort of vortices. In a move Hume would condemn, Newton hinted that at least for planetary motions, immaterial God was the source of the force, but he also hoped that something intermediate between God and gravity could be discovered through experiment.

The two levels of explanation are clear here. The sure method of analysis and synthesis proved the existence of an inverse square force, but nothing explained it, short of God. Newton declared that he had done enough, and that declaration became a guide to the bounds of science. [20, p.45] Hume, however, suggested that Newton had already gone too far in moving from a description of motions to the idea of force.

2.1.4 Force

> the ultimate ontological status of forces in Newton's conception of nature is a complex and involved question. [115, p.377]

Newton struggled with the idea of force throughout his life, changing his ideas several times. Westfall's *Force in Newton's Physics* [115] tracked this development. Notes in Newton's "Waste Book" from 1665 give a version of force strongly influenced by Descartes and based on the impact model. A somewhat later piece, *De gravitatione et æquipondio fluidorum* argued against Descartes that there were absolute motion and space, distinguished from relative motion and relative space by force. "Force is the causal principle of motion and rest." (Newton, *De Gravitatione*, in [115, p.362]) But at this time, Newton had not formulated the relatively simple notion of force that was to characterize his later dynamics. He still used the expression 'force by which a body... indeavours from the center,' an echo of the Cartesian concept of 'conatus'. [20, fn 3][4]

In his 1672–75 correspondence with the Royal Society leading to his optical paper, "An Hypothesis explaining the Properties of Light," Newton espoused a mechanical world view, complete with an æther theory. About five years later he abandoned the æther theory, around the time he was composing "Deære et æthere," which Westfall dated to 1679.

> *De aere et aethere* consists of two chapters. Chapter one, *On Air*, went further than anything Newton had heretofore written in its assertion of forces between particles. Whereas the letter to Boyle had spoken of an 'endeavour to recede,' he now asserted flatly that bodies repel each other at a distance. [115, p.374]

[4] Cohen in turn recommends the reader examine Westfall's *Force*, ch.7.

Although Newton apparently planned to explain that repulsion by means of the æther, the draft of that chapter stopped abruptly, and Westfall speculated that the timing coincided with Newton's revised pendulum experiments which convinced him that there was in fact no material æther.

In 1684 Newton composed "De motu corporum in gyrum," in several versions. The "De Motu" moved towards the concepts of inertia, mass, momentum, and impulse we recognize in Newtonian physics, and parts of the "De Motu" served as drafts for Books I and II of the *Principia*. In 1686, while writing the *Principia*, Newton composed a passage which unambiguously endorsed forces acting at a distance. Originally intended as a conclusion, and then a preface, it was discarded, leaving the guarded statements for which the *Principia* is famous. However, in Query 31 of the first Latin edition of the *Opticks* (1704), Newton inquired,

> Have not the small Particles of Bodies certain Powers, Virtues, or Forces, by which they act at a distance, not only upon the Rays of Light for reflecting, refracting, and inflecting them, but also upon one another for producing a great Part of the Phaenomena of Nature? For it's well known, that Bodies act one upon another by the Attractions of Gravity, Magnetism, and Electricity; and these Instances show the Tenor and Course of Nature, and make it not improbable but that there may be more attractive Powers than these. For Nature is very consonant and conformable to her self. (In [115, p.378].)

Westfall credited Newton with adding "forces of attraction and repulsion... to the catalogue of of nature's ontology," (p.377) thus freeing philosophers from the strictures of the mechanical philosophy, and allowing them to get down to the quantitative work of modern science.

If Newton added forces to the ontology of nature, then how could it also be true that,

> From the point of view of Newton's ultimate metaphysics, then, forces were no more real entities in the universe than they were from the point of view of orthodox mechanical philosophy. [115, p.398]

It is true that Newton thought that gravity was insufficient to have brought the universe from some original state to the ordered system of planets he beheld. Newton thought that an active and immaterial God pervading all of space was necessary to finish his system. Westfall was referring in the above statement to this invocation of a Deity, but his assessment seems too negative.

One could have imagined that Westfall meant "ultimate" in a temporal sense, indicating that while writing Query 31, Newton held a view that forces were real, but ultimately gave them over directly to God. However, that interpretation is hard to reconcile with his claim that, "More fully than any other document, Query 31b embodies his [Newton's] ultimate philosophy of nature." [115, p.379] I think, ultimately, that Westfall disapproved of the mechanical philosophy. In many places he blamed it for slowing down science, and credited Newton with shaking off its oppressive dogma.

> The demand for causal mechanisms had constantly thwarted the drive to express the mathematical regularities in nature. Newton's immaterial aether, the omniscient God, was free of exactly that shortcoming. ... Such a convenient medium delivered him immediately from the preoccupation with causal mechanisms and directed

> his steps down the royal road toward a quantitative, as opposed to a verbal, dynamics. (pp.398-9)

There is a way to see in Newton both a drive towards causal explanation and one towards the mathematical expression of regularities.

Early in his scientific career, Newton had moved beyond his youthful dependence on the æther theory and, however privately, decided that forces really were the ultimate *physical* levels of explanation. However, he searched for another level of explanation. Having found no material qualities which would explain all the forces including gravity, and perhaps thinking that in Book II of the *Principia* he had removed the possibility of *any* material æther, he looked to an immaterial æther, a Deity who could supply the active power needed for attractions and repulsions.

> For two planets separated from each other by a great expanse of void do not mutually attract each other by any force of gravity or act on each other in any way except by the mediation of some active principle that stands between them by means of which force is propagated from one to the other. . . . Therefore the ancients who grasped the mystical philosophy more correctly taught that a certain infinite spirit pervades all space, and contains and vivifies the entire world; and this supreme spirit was their numen; according to the poet cited by the Apostle: In him we live and move and have our being. Hence the omnipresent God is recognized. . . . (*Add.MS* 3965.6, f.269, in [115, pp.397-8])

On this view, gravity was the ultimate court of appeal for natural philosophy, and beyond that there lay only God. God was the cause of the instantaneous action at a distance Newton found so hard to reconcile with natural philosophy. Yet, convinced that he was right in his *description* of gravity, Newton concluded that he could infer directly to an omnipresent, omnipotent deity. Indeed, it is hard to see what else could fit.

Cohen's reading, on the other hand, allowed a span of causes between the raw assertion of forces and the action of the Deity. This makes sense of Newton's Queries on elastic and electric ethers, and his method of proceeding "from effects to their causes and from particular causes to more general ones, till the argument end in the most general." In the end, the most general was always God, but the Queries seem more hopeful that philosophy had not reached the limit where God must be invoked directly. The difference may well have been that the causes needed to explain optical and chemical phenomena were obviously short-range forces, where the infinite range of gravity was so clearly beyond the scope of natural philosophy.

Although modern science is still very Newtonian, nothing of Newton's natural theology remains in it. Hume's critiques played a large role in removing the one from the other. Hume objected that it was not possible to infer from physical principles to an omnipotent Deity, and that the proper attitude was to admit that with a formulation of universal laws of motion, reason could go no farther. However, Hume's quarrel was not, as he saw it, so much with Newton as with Newton's followers and critics.

2.2 Overview: Newton to Hume

The eighteenth and nineteenth century saw natural philosophers and scientists turn away from inquiries into mechanical causes and towards mathematical, functional relationships between quantities. With the benefit of two centuries of scholarship and the gradual discovery of more of Newton's works, we know now a great deal more about Newton's views than his eighteenth-century followers did. Very few of Newton's contemporaries suspected his true religious sympathies, and neither were Newton's extensive theological and alchemical researches known or available to the eighteenth-century Newtonians. Instead, they had Newton's published works, several reviews and simplifications, and a handful of laudatory but somewhat inaccurate biographical essays.

Yet Newton's success was unparalleled, and his foundations and methods, or related ones, had already proven valuable in furthering celestial and fluid dynamics. The Newtonian scientific method and world view spread rapidly, but not without assistance. Newton's methods, aims, and achievements had already been misunderstood in England because they did not square with the reigning Cartesian rationalist philosophy. John Locke undertook to clear away rationalist misconceptions about knowledge and justify Newton's empirical approach.

In carrying out that plan, Locke unwittingly set the stage for Hume's demolition of key pieces of empiricism, limiting the scope of knowledge more severely than anticipated. Because we know that Hume read and reacted to Locke, and because Locke's work became so tightly bound to Newton's, especially on the continent, we turn next to that philosopher's interpretation of Newton.

2.3 Locke (1632-1704)

> He that, in the ordinary affairs of life, would admit of nothing but direct plain demonstration, would be sure of nothing in this world, but of perishing quickly. *EHU,* IV.xi.10

Locke had been working on his *Essay Concerning Human Understanding* before the *Principia* was published, and his longstanding friendships with Boyle and other members of the Royal Society may have been at least as significant in determining his affinity for empiricism. So may his own pursuits in medicine and chemistry. However, Locke read the *Principia* soon after it was published, wrote one of its first reviews, and soon afterwards became friends with Newton. Although their correspondence was mostly on religious and chemical matters, their philosophies were consonant enough to become fused in the minds of eighteenth-century philosophers. Said philosophers took at face value Locke's claim to have been merely an "Under-Labourer . . . removing some of the Rubbish, that lies in the way of Knowledge." Indeed, before 1700 Bentley, Whiston, Molyneaux, and Leibniz had woven several Lockean themes into their conception of Newtonian philosophy, and Locke had preceded Hume in insisting that the science of man was the foundation for all other knowledge. [33, pp.296-7] The Newton which Hume knew had already been amalgamated with Locke's empiricism, and had spread throughout almost all areas of popular culture.

Locke studied medicine and chemistry at Oxford, and after reading Descartes turned to philosophy. Although he credited Descartes for being clear and comprehensible, Locke opposed Cartesian rationalism. As an empiricist, he denied that there were *any* innate ideas, undermining Descartes' top-down approach to natural philosophy. Locke used the word "ideas" for many kinds of mental phenomena, but his empiricist stance claimed mainly that there were no innate sensations or reflections. All such ideas came ultimately from experience with the world. In addition, he was a representationalist; he claimed that people perceived their ideas of external objects, not the objects themselves. However, those ideas *represented* external objects because those objects had qualities which produced ideas in the mind. Berkeley took the next step and claimed that people had access *only* to their ideas, and that talk of external objects was already bad metaphysics.

Locke however, had not completely abandoned the Cartesian world picture. For instance, he separated primary and secondary qualities. Primary qualities like extension "really do exist in the bodies themselves", although we do not perceive them, while secondary (perceived) qualities like color were in the mind, but not the object.[5] However, Locke's corpuscularian leanings were tempered, and in the end he argued that we could not really know what the true primary qualities were. [27] Locke's empiricism and his theory of representation were, while perhaps not as heavily influenced by Newton and Boyle as once thought, certainly of a piece with the whole approach to natural philosophy engendered in the British Royal Society. [93] In his *Essay* (1689), and especially Book IV of that work, Locke argued that knowledge of the physical, sensible world could only come from inductive, experiential knowledge, and not from rational introspection. This was a very Newtonian theme, and one which Hume would take up and elaborate upon. In subsequent revisions, Locke incorporated more references to Newton, adapting his own philosophy slightly to accommodate gravity as a principle of the world. [93]

Unlike Berkeley, who wrote in direct response to Locke, Locke himself did not focus so clearly on a particular opponent so much as a general dissatisfaction with the existing state of philosophy, which in his experience had been largely scholastic. Locke created a vast sprawling system which allowed for empirical *knowledge* as well as *belief*, breaking the Platonic veil of perception. In establishing the "extent of *human knowledge*" [73, I.1.2], Locke explicitly sought to understand the capacities and limits of knowledge in natural philosophy, especially as practiced by Newton and Boyle. This was also Hume's project, and we see in Locke the ideas against which Hume reacted.

We are most concerned with Locke's thoughts on causality, so let us begin with Book II, "Of ideas." In Book II, Chapter XXV, "Of relation," Locke declared that "relations all terminate in simple ideas," including cause and effect.

> I shall begin with the most comprehensive relation, wherein all things that do, or can exist, are concerned, and that is the relation of *cause* and *effect*; the idea whereof, how derived from the two fountains of all our knowledge, sensation and reflection, I shall in the next place consider. [73, p.203]

Chapter XXVI, "Of Cause and effect, and other relations" followed immediately. We learn there that "That which produces any simple or complex *idea* we denote by the general name, *cause*, and that which is produced, *effect*." This is, of course, circular in that it assumes the reader can define "produce."

After a small discussion on the differences between creation and alteration, Locke gave a more useful description of causation, or at least how it was and how it was grounded in sense data, that is, when something is perceived to cause something else. [73, p.205]

> In which, and all other cases, we may observe that the notion of cause and effect has its rise from ideas received by sensation or reflection; and that this relation, how comprehensive soever, terminates at last in them. For to have the idea of cause and effect it suffices to consider any simple idea or substance, as beginning to exist, by the operation of some other, without knowing the manner of that operation.

[5] D. Bertoloni-Meli has pointed out to me that Newton did not share this view.

The last clause is particularly relevant. In a very Newtonian manner, it makes causation an empirical phenomenon. Mechanist philosophers would have demanded some sort of mechanical account of how effects were brought about, but Locke held that this was not necessary, opening the way for an empirical, mathematical philosophy which put aside ultimate questions. This move was important philosophically, since Newtonians had to argue that a theory which apparently failed to provide an acceptable causal explanation could still count as "an acceptable piece of natural philosophy." [27, p.294]

Locke's account so far was incomplete in that he had not really described how to decide when the existence of an idea or substance was due to "the operation of some other." As we saw in a previous quotation, Locke had a "production" model of causality. Hume would complain of precisely this kind of circular definition, and offer his own more empirical one. According to Hume, causation corresponded to observed correlation or covariation, plus a habit of the mind to make an association between correlated or covaried ideas. This radical move is not without its problems, but at any rate Locke does not seem to have thought of it.

Furthermore, despite his efforts to make room for empirical knowledge, it is not clear where in Locke's typology (see table 2.1) Newtonian theories really fit. Terrestrial gravity was an empirical phenomenon, and fell under knowledge of coexistence. Universal gravity was something else altogether. It arose from Newton's use of "mixed mathematics," the application of the demonstrative knowledge of mathematics to the probable knowledge of the empirical world. Newton claimed for it a kind of certainty, which was not the same certainty that could be achieved among pure relations of ideas such as logic or mathematics. Like Hume, Locke did not appreciate the role that quantitative, mathematical analysis played in the new science.

Kind	**Degree**	**Example**
Identity, or diversity	Intuitive	"Blue is not yellow."
Relation	Demonstrative (Certain but not obvious)	"Two triangles upon equal bases between two parallels are equal."
Coexistence or necessary connexion	Sensitive (Empirical, tentative)	"Iron is susceptible of magnetical impressions."
Real existence	Categorical Intuitive (own existence) Demonstrative (God's) Sensitive (anything else)	"God is."

Table 2.1: Locke's four kinds of knowledge (Book IV).

2.4 Berkeley (1685-1753)

> Allot to each science its own province. *De Motu*, p.72

Not all theologians and philosophers approved of Newton's and Locke's natural philosophy and natural theology. Berkeley argued against Newtonian ideas of absolute space, time, and motion, against corpuscularianism, and against Locke.

Berkeley held that such unobservable terms as Space, Time, Force, Gravity, and Attraction were merely conceptual devices. Although they were useful shorthand for working scientists, they did not correspond to anything in reality. According to Berkeley, there was no efficient cause but spirit, which was beyond the scope of natural philosophy. Because Berkeley's idealism is so extreme, it is easy to forget that his objections reflected a raging controversy over the interpretation and foundations of the calculus and its place in natural science, and that his was a genuine attempt to cut through the metaphysical tangles introduced by Locke's philosophy. In particular, he attacked the theological conclusions reached by Newton and his followers, especially Clarke. Although Hume was no idealist, he borrowed Berkeley's approach in pushing Locke's theories to their ultimate empirical conclusions.

We are most interested in Berkeley's *A Treatise Concerning the Principles of Human Knowledge*, especially sections 97-8, and 101-117, and his *De Motu*. Berkeley thought that philosophers made things too hard for themselves by concentrating on abstract ideas like time, extension, and motion. For Berkeley, there was no abstract, colorless extension, only extended things.[6] Likewise, he thought that it was more sensible to dispense with "true natures" altogether, rather than argue that they were hidden from the senses. This assumption of unknowable essences handicapped the study of nature.

Among internal essences or principles, Berkeley singled out attraction as a non-explanatory non-entity.

> The great mechanical principle now in vogue is *attraction*. That a stone falls to the earth, or the sea swells towards the moon, may to some appear sufficiently explained thereby. But how are we enlightened by being told this is done by attraction? . . . I do not perceive that anything is signified besides the effect itself; for as to the manner of the action whereby it is produced, or the cause which produces it, these are not so much as aimed at. [5, *103*]

In fact, Berkeley thought it wrong to presume that gravity was universal. Although "gravitation or mutual attraction" appeared to be common, he denied that it was universal or essential. His arguments were not very good. For example, that "it appears the fixed stars have no such tendency towards each other," and "the perpendicular growth of plants" indicated to him a lack of universal attraction. (*106*) All of this was in keeping with his idea of the variety of the "Governing Spirit," and perhaps also with an inappreciation of the mathematics of universal gravitation. His general philosophical point was much more persuasive than the evidence he offered in its favor. Hume would strengthen the skeptical background behind the denial of an inference to universals. And indeed, before bringing in the nature of the force, what did "attraction" add to the laws of motion, except that it was a term in the mathematical account of the motions.

Newton had allowed that ultimately, God was the cause of everything. Berkeley, however, upped the ante. Since everything depended on the will of the Governing Spirit,

> it is plain philosophers amuse themselves in vain, when they enquire for any natural efficient cause, distinct from a *mind* or *spirit*. Secondly, considering the whole creation is the workmanship of a *wise and good Agent*, it should seem to become philosophers to employ

[6] Compare this to James Gibson's ecological psychology, which is normally quite critical of Berkeley.

> their thoughts (contrary to what some hold) about the final causes of things...the various ends to which natural things are adapted, and for which they were originally with unspeakable wisdom contrived.... (*107*)

Intermediate, physical causes even as abstract as Newton's forces were removed. This did not for Berkeley imply the end of empirical science, for if God acted uniformly, one could discover the laws of nature and make predictions. Berkeley thought that nothing essential to science would be lost. However, the science Berkeley had in mind was not Newtonian, since he compared the mathematical approach to paying attention to the "grammar-rules" and signs rather than the causes. (*108*) Unlike Locke and Hume, Berkeley was actively engaged in the contemporary metamathematical debates. However, like Locke and Hume, Berkeley had no real use for mathematics in natural philosophy, at least not in an essential way. If we recall that the mathematics of fluxions and infinite series had no good foundations at that time, this is perhaps excusable.

In addition, Berkeley strongly disagreed with Newton's stance on absolute time, space, and motion. Imagining a universe with only one body, he argued that in such a case no motion would be conceivable. Similarly, Berkeley defined space as the absence of body, definable only with respect to bodies, as regions of unimpeded motion. To ascribe more to it than that was to make an error of reification. Berkeley was strongly motivated to make this point, since otherwise one could be led to think that "Real Space is God, or else that there is something beside God which is eternal, uncreated, infinite, indivisible, immutable," which was unacceptable. *(110)*

Berkeley denied the idea of number in the abstract, thought that the study of theorems was a useless devotion to signs instead of things, and held that extension was not infinitely divisible because we do not perceive an infinite number of parts. In an apparent reference to Newton's predecessor Isaac Barrow, whose mathematical lectures had been published around 1670, Berkeley consoled those who feared that thereby "the very foundations of Geometry are destroyed," by noting that at a practical level,"whatever is useful in Geometry, and promotes the benefit of human life, does still remain firm and unshaken on our Principles." *(117)* The smallest Berkeley would allow was the "*minimum sensible*", an idea Hume also adopted.

Later, in his 1721 *De Motu*, Berkeley was more direct. As he put it, the main point was that "In the pursuit of truth we must beware of being misled by terms which we do not rightly understand." *(1)* Chief among these were 'solicitation,' 'effort,' 'conation,' 'force,' and 'gravity.' The first three "belong properly to animate beings alone," while 'force' and 'gravity' were occult qualities and needed to be discarded except in reference to particular bodies and motions. *(3,4,5,6)*

A section in Berkeley about the relationship between force and momentum is important for showing how Newton's ideas were understood at the time.

> The force of gravitation is not to be separated from momentum; but there is no momentum without velocity, since it is mass multiplied by velocity; *(11)*

We shall later look at a section where Hume seems to have a confused notion of force. In fact, however, it is this same conjunction between force and momentum, a reading that was well within the normal contemporary understanding of Newton, as we shall see in a moment when we turn to Colin Maclaurin. That understanding was, of course, influenced by pre-Newtonian ideas of force, as illustrated by Berkeley's discussion of measuring forces by their impacts and impressions in soft materials.

At any rate, eighteenth-century philosophers were not unanimous in their interpretation of force. Berkeley gleefully observed,

> Leibniz confuses impetus with motion. According to Newton impetus is in fact the same as the force of inertia. Borelli asserts that impetus is only the degree of velocity. Some would make impetus and effort different, others identical. Most regard the motive force as proportion to the motion; but a few prefer to suppose some other force besides the motive, to be measured differently, for instance by the squares of the velocities into the masses. *(16)*

Thirty years later, Hume would still have trouble sorting out the issue of force, and end up with empiricist conclusions not far from Berkeley's. Berkeley saw the reification of abstractions as the source of the problem. He noted in his own defense that Newton introduced attraction, "not as a true, physical quality, but only as a mathematical hypothesis," and that Leibniz likewise considered effort and impetus to be abstractions. *(17)*

The power of moving bodies did not lie in the bodies themselves, in some innate principle of attraction or repulsion, but in soul. Our own experience showed us that "bodies are moved at the will of the mind." *(25)* Even "Newton everywhere frankly intimates that not only did motion originate from God, but that still the mundane system is moved by the same actus." *(32)* Here Berkeley and Hume diverged, for Hume thought that if we can not infer innate forces, still farther from experience was God himself.

Moving closer to the billiards table, Berkeley took up the "cause of the communication of motions," *(67)* arriving at a double conclusion. First, God is the cause:

> That the Mind which moves and contains this universal, bodily mass, and is the true efficient cause of motion, is the same cause, properly and strictly speaking, of the communication thereof I would not deny. *(69)*

However, in natural philosophy

> we must seek the causes and solutions of phenomena among mechanical principles. Physically, therefore, a thing is explained not by assigning its truly active and incorporeal cause, but by showing its connection with mechanical principles, such as *action and reaction are always opposite and equal.* From such laws as from the source and primary principle, those rules for the communication of motions are drawn, which by the moderns for the great good of the sciences have been already found and demonstrated. *(69)*

Properly divided thus, Berkeley allowed the mechanical philosophers to talk as if the striking body causes the motion of the struck body. However, he asserted that this was just for convenience of speech and the further development of the rules of motion, but that we should understand that the true efficient cause is an active principle of God.

Although Newton would almost certainly have disagreed with Berkeley's assessment of the role of mathematics in natural philosophy, and his treatment of infinitesimals and infinite divisibility, this separation of levels was in keeping with Newton's own two-level distinction between causes of motion (forces) and the causes of those causes (usually, God). The difference was that Newton thought forces were real, physical entities, where Berkeley thought they were useful mathematical fictions and shorthand.

One of the most vocal defenders of Newton against Berkeley was Colin Maclaurin, a mathematician at Edinburgh from 1725 until his death in 1746. Maclaurin was an able Newtonian and an ardent proponent of Newtonian natural theology. He considered the inferences to cause, effect, mind, spirit, and God to be completely transparent and unproblematic, and defended the calculus and its attendant notion of infinite divisibility. Hume almost certainly read Maclaurin, and may even have studied under him while at Edinburgh. Nevertheless, as we shall see while developing the details of Hume's position, there was little agreement. First, we shall set out Hume's fundamental account of causation.

2.5 Hume (1711-1776)

David Hume grew up in a philosophical landscape dominated by Newton's philosophy, which he self-consciously imitated, hoping at least early in his career to create a "science of human nature" to parallel Newton's achievement in mechanics. Hume was probably still attending the University of Edinburgh when the third edition of Newton's major work, the *Principia*, was published in 1726. This third edition contained the fourth rule of reasoning which Newton had added to avoid certain criticisms, and which Hume later turned back against the Newtonians to curb what he saw as their theological excesses in natural philosophy.

Hume entered the University of Edinburgh around 1723 or 1724 and after finishing attempted to study law but instead spent his time on "philosophy and general learning." In 1734 (age 23) he turned for a short and unproductive time to business. Rejecting that path, he decided to pursue a life of learning and went to France where over the next three years he wrote his first and major work, *A Treatise on Human Nature*, which was famously unsuccessful. However, after some other diversions his essays became well known, and he gradually reworked his treatise, publishing the *Inquiry Concerning Human Understanding* in 1748, and *Inquiry Concerning the Principles of Morals* in 1751. [55, "My Life"] At that time he began writing his *Dialogues Concerning Natural Religion*, which he then shelved until just before his death in 1776.

2.5.1 Overview of Hume's causal account

The skeptical attack

Hume argued first that causes were not knowable *a priori*, that is, by reasoning alone. If a cause were known to reasoning alone, then the effect E must follow deductively from the cause C, but no effects are deductive consequences of their causes. For any given cause C, we can imagine or conceive any effect, E. It implies no contradiction to imagine that a dropped glass would remain in place or fall up. Likewise, it is logically possible for a cue ball to pass through the eight ball instead of colliding with it. We know by experience that this never happens, but the link between C and E is not deductive, and therefore not knowable to reasoning alone. Another way of saying this is that there is nothing in the cause which logically entails the effect.

The constructive account

But if observation and experience do not yield a sense of necessary connection, from where do we gather our idea of a causality? If no cause implies its effect, how do we know what to expect? According to Hume, the observation of regularities in the world provides provides the basis for our impression of causality. In fact, the "necessary connection" was merely the expectation of the effect—the *sense* of a necessary connection—built up by repeated experience with a world that happens to act regularly. Our minds are so disposed as to register this regularity, moving by force of habit from repeated experience to expectation. Why that happened to be so Hume was not able to say, and here he took a very Newtonian line. Certain ideas in the mind attract each other and cohere, but Hume could not claim to know the nature of that attraction.

Implications

Causal links are psychological additions to a world sufficiently governed by the laws of motion. Hume does not doubt that we have an impression of causality, but only that there is some ultimate rational justification for our habitual "inference" from cause to effect. That inference or expectation remains for Hume an empirical induction from repeated experience. All we really perceive are constant conjunctions of objects or events, to which we attune. When we see two events regularly conjoined in space and time, the first or prior one naturally and forcefully brings to mind the second or consequent one. Hume accepts this as a brute fact about the action of the mind, and calls the resulting forceful expectation our causal impression, much as Newton called the universal attraction a "centripetal force," which was meant to imply no more than that there was a force directed toward the center.

It is fortunate that we have such a faculty for causal impressions, said Hume. This faculty allows us to make judgments which are often correct, and yet which reason could never reach, because they are not rationally justified. As he wrote in his *Abstract:*

> It is not, therefore, reason which is the guide of life, but custom. That alone determines the mind in all instances to suppose the future conformable to the past. However easy this step may seem, reason would never, to all eternity, be able to make it. [54, p.189]

Like Locke, Hume concluded that had we been left with reason alone we would have fared poorly, because "It is evident that all reasonings concerning *matter of fact* are founded on the relation of cause and effect. . . ." [54, p.186] Knowledge of causes was based only on experience, and all knowledge of the world was based on knowledge of cause and effect.

2.5.2 Definitions

It was not easy to pin down this idea of causality. Notoriously, Hume offered several definitions of causality which do not necessarily square with one another. There are three basic varieties, the regularity, psychological, and counterfactual definitions. They are given in the *Treatise* and in the *Inquiry*.

Regularity, or constant conjunction

In the *Treatise,* Hume wrote:

> We may define a CAUSE to be an object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedence and contiguity to those objects that resemble the latter. [53, p.170]

and later in the *Inquiry*, he wrote:

> we may define a cause to be *an object followed by another, and where all the objects, similar to first, are followed by objects similar to the second.* [55, p.87]

The first of these exhibits Hume's classic triad: priority, contiguity, constant conjunction, although the second has only two of these. We will review this difference shortly. Here and elsewhere, where Hume writes "objects," I think we have to read "events," since even his famous billiards examples refer to the cause of motion, not the cause of the ball. The account loses coherence if we require Hume to mean what we would call an object. That difficulty reconciled, the regularity account makes the strong claim that any constant conjunction is a cause, where a constant conjunction is a perfect correlation between two events which are adjacent in space and time. This is a metaphysical rather than an epistemological notion of cause since it is objective, at least so far as resemblance is objective.

Psychological, or expectation

From the *Treatise,*

> A CAUSE is an object precedent and continuous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other. [53, p.170]

From the *Inquiry,*

> *an object followed by another, and whose appearance always conveys the thought to that other* [55, p.87]

This definition is much more in the spirit of Hume's empiricist program. The notion of a cause is native to psychology, not physics. Whatever the case out in the world, a cause is anything which brings about certain impressions in (all?) observers.

Counterfactual

The counterfactual definition immediately follows the regularity definition in the *Inquiry.* Here I have left only the counterfactual definition in italics:

> where all the objects, similar to first, are followed by objects similar to the second. Or, in other words, *where, if the first object had not been, the second never had existed.* [55, p.87]

The pretended equivalence of these two sentences does not withstand scrutiny. The counterfactual account is much stronger since it invokes the very sense of necessity that Hume sought to explain away. The difference between sentence one (constant conjunction) and sentence two (counterfactual) is the difference between universal generalization and natural law [28]. The inequivalence between the two is the source of difficulty for empiricist accounts from Hume onwards. Accidental universal generalizations[7] and lawful universal generalizations[8] both have the same *lawlike* form, but one supports counterfactuals[9] while the other does not.[10]

[7] "All diamonds have a mass less than 500kg."

[8] "All diamonds are made of carbon."

[9] "If this massive boulder were a diamond, it would be made of carbon."

[10] "If this massive boulder were a diamond, it would have a mass less than 500kg."

Discussion

There is an interesting change in emphasis between the *Treatise* and the *Inquiry*. Both definitions in the *Treatise* stipulate contiguity, yet all three definitions in the *Inquiry* omit it. In the *Treatise*, Hume wrote,

> I find in the first place, that whatever objects are consider'd as causes or effects, are *contiguous;* and that nothing can operate in a time or place, which is ever so little remov'd from those of its existence. Tho' distant objects may sometimes seem productive of each other, they are commonly found upon examination to be link'd by a chain of causes, which are contiguous among themselves, and to the distant objects; and when in any particular instance we cannot discover this connexion, we still presume it to exist. [53, p.75]

As it turns out, Hume never fully established the essential role of contiguity in the *Treatise*, and perhaps consequently, never mentioned contiguity as a part of causality *anywhere* in the *Inquiry*. [107, pp.43-44]

As Hume admitted in the preceding quotation, and has been born out in experiments on perception, sometimes we have the impression of causation without an impression of contiguity. Considering the time at which he wrote, gravity and magnetism were probably foremost on his mind. In fact, it is tempting to see here a replay of Locke: an early insistence on causation by contact, and a gradual relinquishing of that view as the author comprehended the success of Newton's gravitational theory.

We can also have impressions of contiguity, even constant contiguity, without receiving a causal impression. This is the case with many universal generalizations which we know are not causal laws, such as night and day, or any number of accidental generalizations. [107, p.66] Indeed, learning experiments in animals and humans have demonstrated that under certain conditions learners will completely fail to attend to perfect correlations if they already think that some other variable is more predictive. This is the well-known "blocking" effect which was used to attack conditioning theory. For example, a dog is conditioned to salivate when it hears a bell, by always following the bell with meat powder. Then, another conditioning trial is undertaken where the bell is always accompanied by a light. However, when presented with only the light, the dog fails to salivate, despite the excellent correlation between light and reward. The previous association between the bell and the reward is said to have blocked the learning of the new association.

However, these limitations to Hume's definition do not hinder his main skeptical argument. Even if we had priority and contiguity, that would not be enough. "Two objects might be related by contiguity and priority in time 'merely coincidentally'." [107, p.44] I shall interpret Hume as having expressed a strong preference for contiguity and priority, but without absolutely requiring them. Hume wanted to figure out what distinguished mere coincidence from actual causation, and the only thing he saw, short of invoking notions of power, force, and energy, were his principle of constant (or for-the-most-part) conjunction. We have already seen some shortcomings of this position. In the next chapter we will examine modern attempts to address those shortcomings.

2.5.3 Hume's use of Newton

> The claim that Hume's philosophical development was profoundly affected by the Newtonian method is as probable as any assertion of an unacknowledged influence can be. [87, p.76]

Hume's sources

The landscape for those studying Hume's life is somewhat bleaker than the lush foliage facing the Newton scholar. Hume's brief autobiography tells us little about his influences, and his letters mention neither Newton (except once, incidentally and facetiously), nor other philosophers with whom he is known to have been familiar. [87, p.75] Aside from references in the texts, the best we have to go on is a letter from Hume to his friend Michael Ramsay, about whom little is known [11, p.11], in 1737. The latter had asked Hume what books would provide an appropriate background for reading the *Treatise*. Hume recommended the following: [35]:

- Bayle's Dictionary: entries on Spinoza, and Zeno[11]
- Malebranche: *Recherche de la Vérité*
- Descartes: *Meditations*
- Berkeley: *Principles*

Based on references in the *Treatise* itself and work done by Kemp Smith comparing Hume's passages with potential sources, Frasca-Spada has added the following [35]:

- Newtonian theology, *eg* Clarke
- Scottish moralists, *eg* Hutcheson
- Locke: *Essay*, and *L'Art de Penser*
- Barrow: *Usefulness of Mathematical Learning*
- Malezieu: *Elémens de Géometrie*

That Hume was familiar with Newton is clear, but we do not know how much he read directly. The whole program of extending the experimental method to the moral sciences sounds very Newtonian—compare the following from Newton's *Opticks*:

> And if natural philosophy in all its parts, by pursuing this method, shall at length be perfected,the bounds of moral philosophy will be also enlarged. (Question 31, in [108, p.179])

However, the only direct *citation* of Newton occurs in the *Enquiry Concerning the Principles of Morals* in a footnote at the end of section III [57, p.38]. The selection reads, including Hume's own footnote,

> It is entirely agreeable to the rules of philosophy, and even of common reason; where any principle has been found to have a great force and energy in one instance, to ascribe to it a like energy in all similar instances. This indeed is Newton's chief rule of philosophizing.[a]
>
> [a]Principia, Lib.iii.

[11] Of possible significance: Bayle's dictionary at the time did not have an entry on Newton.

We should note two things about this passage. First, this appears to be a reference to Newton's Rule II, "Therefore to the same natural effects we must, as far as possible, assign the same causes." Second, that Hume has admitted that such a rule is "entirely agreeable" to natural philosophy, even if its ultimate justification is unavailable.

There are, as we shall see, other places where Hume makes reference to Newton, although it is harder in those cases to tell whether Hume was getting his Newton direct from the source or through Locke, Berkeley, Maclaurin, or others. We have already mentioned the possibility that Hume had met Maclaurin or even attended his lectures while at Edinburgh. Even before Maclaurin however, Edinburgh was a stronghold of Newtonian thought. It was, outside of Cambridge of course, the foremost center of Newtonianism in England, and had begun to adopt Newtonian ideas even before 1700. [102]. Kemp Smith [103, p.53,n.1] noted that Robert Stewart, the Chair of Natural Philosophy at Edinburgh from 1708 to 1742,

> is reported to have been in his earlier years a Cartesian, and later a Newtonian. Hume probably attended his class in the Session 1724-5. What Stewart's teaching then was, we can only conjecture.

Adding to the circumstantial evidence, Hume said he conceived his idea of an experimental philosophy of the moral sciences while he was of university age. Several biographers have speculated that it was Hume's contact with Newtonian thought which sparked this idea, and in an early letter Hume did mention a "new scene of thought," which sounds like a reference to the Newtonian philosophy.

Be that as it may, it is almost certain that Hume read Maclaurin in his later life. Maclaurin's *Account* was published in 1748, the same year as Hume's *Inquiry*. The lengthy List of Subscribers published with the book included "The Right Hon. the Earl of Home" who is marked as one who "subscribed for the large paper." Since Hume mentioned in his autobiography

> My father's family is a branch of the Earl of Home's or Hume's; and my ancestors had been proprietors of the estate, which my brother possesses, for several generations. [55, p.3]

I thought it was possible that this subscription entry referred to Hume's brother John Home, with whom Hume lived from 1742 to 1751 except for two years (including the first part of 1748) of work as a tutor and aide-de-camp. Fortunately for us, David Hume's library and his brother's were both passed on to the philosopher's favorite nephew, also named David, who became the Baron Hume, and whose will called for an inventory of possessions. That inventory included a 1748 quarto edition of Maclaurin's book. [86, p.111] The timing is suggestive. Hume returned to England in 1748 and in that year came out with the *Philosophical Essays*, which is to say the first edition of the *Enquiry Concerning Human Understanding*.

Hume, Maclaurin (1698-1746), and Natural Philosophy

There is already good evidence that Hume lifted a passage for his *Dialogues*[12] straight from Maclaurin [87, pp.102-104]. I would like to argue here for the possibility that Maclaurin influenced the *Inquiry* as well, despite publishing in the same year. First, Maclaurin began to write his work quite early and may have finished it as early as 1728, even though it was published only posthumously. Certainly he was giving Newtonian lectures before that. Second, Maclaurin had

[12] Begun around 1750, finished in 1776 and published posthumously.

already in 1742 published some of his views in response to Berkeley's 1734 *Analyst*. Third, although Hume apparently possessed a copy of Maclaurin, there is no record that he owned a copy of Newton, although his conviction that he knows Newton better than the Newtonians at least suggests that Hume had somewhere read Newton himself. Fifth, in the one section of the *Inquiry* where Hume explicitly addressed natural philosophy, his argument resembled part of that given in Maclaurin.

Hume did not often discuss natural philosophy in any direct way, which has led to the impression that he had only a passing or at least a nontechnical understanding of Newton's work.[13] For example, even in his book defending the Newtonian character of Hume's thought, Noxon wrote,"I doubt that Hume had mathematics enough to read the *Principia*, if 'reading' includes following the geometrical demonstrations." [87, p.69] Similarly, after pointing out eleven explicit references to Newton in Hume's writings,[14] Force noted that they indicated only an interest in "Newton's methodological principles," not his "mathematics and mechanics." [34].

Apparently Hume knew enough geometry to attempt a dissertation on the subject, but its fate can hardly be said to endorse the author's grasp of the material.

> it is also known that, around 1755, Hume wrote a dissertation on geometry. Lord Stanhope 'convinced me, that either there was some Defect in the Argument or in its Perspicuity: I forget which'. . . . at any rate, the dissertation, which was meant to appear in a collection published in 1757 and called *Four Dissertations*, was suppressed. [36, p.96]

In fact, much of Frasca-Spada's book was devoted to untangling the apparent mess Hume made of geometry and other mathematics in his discussion of the limits of perception and infinite divisibility. Frasca-Spada traced the arguments in some detail, and while not absolving Hume of error, pointed out that his error was not greater than that of his sources, and that it was largely confined to the contentious area of infinites and infinite divisibility. Hume disagreed with Newton and Barrow, but so did several of Hume's sources such as the great encylopædist Bayle, whose entry on Zeno contained just those crude atomistic dismissals of infinite divisibility that Barrow had shown to be fallacious.

Be that as it may, any use of Newton almost compelled a reassessment of natural philosophy. Locke had been forced to recant his prior position that action was communicated solely by contact. Perhaps his growing familiarity with Newton also explained why Hume had dropped his insistence on contiguity of cause and effect between the *Treatise* and the *Inquiry*? It may have been merely dissatisfaction with his own arguments, but the parallel with Locke is suggestive. Fortunately, Hume's reputation as the Newton of the moral sciences is based on somewhat stronger evidence. Newton emphasized experiment, downplayed speculation, and advocated the idea of universal attraction, whatever it was. In summarizing scholarship on the ties between Newton and Hume, Casini offered the following points of comparison [14, pp.46-47]:

- In the introduction to the *Treatise*, Hume "insisted on the need for universal principles, for the limitation of human knowledge to experience, and for the practice of 'careful and exact experiments'."

[13] It has led some to assert that Hume was completely unfamiliar with and uninterested in the Newtonian system. See [34] for references as well as an argument against that position.

[14] These are reproduced in the appendix.

- In the same, "He forcibly rejected all 'presumptuous and chimerical hypotheses' from the domain of psychological inquiry. . . ."
- "Hume considered attraction as the clue to psychology and morals. He spoke of a 'gentle force' uniting the ideas, and defined this principle of association as 'a kind of attraction, which in the mental world will be found to have as extraordinary effects as in the natural. . . '."
- Hume's eight rules of causal inference display an "unmistakable echoing of Newton's *Rules of philosophizing.*"
- In the *Inquiry,* Hume wrote, "But as to the causes of these general causes, we should in vain attempt their discovery," echoing Newton's own distinction.

For the most part, Hume avoided the subject of natural philosophy, which is why it is so hard to pin down his apparent Newtonianism. We have already noted that, within natural philosophy, Hume allowed that it was customary and proper to use principles of reasoning for which there was not ultimate justification, in particular the rule of "same cause, same effect." As is well known, in his arguments against miracles in the *Enquiry*, Hume's skepticism about the uniformity of nature was nowhere to be seen. But what were Hume's particular thoughts about topics in natural philosophy?

Hume's brief discussion of the laws of motion in the *Enquiry* seems at first to confirm the notion that he did not understand or know about Newtonian mechanics.

> Thus it is a law of motion, discovered by experience, that the moment or force of any body in motion is in the compound ratio or proportion of its solid contents and its velocity, and consequently, that a small force may remove the greatest obstacle or raise the greatest weight if by any contrivance or machinery we can increase the velocity of that force so as to make it an overmatch for its antagonist. (*Inquiry*, p.45 [55])

I will first describe what appear to be the problems with this paragraph, and then explore what it would have meant to eighteenth-century philosophers.

The passage can be split into sections.

> Thus it is a law of motion, discovered by experience, (A)

> that the **moment** or **force** of any **body in motion** is in the compound ratio or proportion of its **solid contents** and its **velocity**, and consequently, (B)

> that a **small force** may remove the greatest obstacle or raise the greatest **weight** if by any contrivance or machinery we can increase the **velocity** of **that force** so as to make it an overmatch for its antagonist. (C)

Section A states a core Humean theme: knowledge comes originally from experience, not reason. Later in the same paragraph Hume explained,

> the discovery of the law itself is owing merely to experience; and all the abstract reasonings in the world could never lead us one step toward the knowledge of it (p.46)

Section B is more troublesome. If Hume was summarizing Newton, he seems to have been confused, or at least imprecise. It seems more likely that he is using a more Cartesian scheme which had not entirely separated dynamics from statics.

In section B Hume asserted

- moment $\propto$ (solid contents)$\times$(velocity)
- force $\propto$ (solid contents)$\times$(velocity)

and has implied that moment and force are synonymous. Although this strikes one as an error, we find the same statement in Maclaurin. [75, pp.105-106],

> 12. The quantity of motion in a body being the sum of the motions of its parts, is in the compounded ratio of its quantity of matter and of the velocity of the motion." (p.105)

For convenience, we can write that the quantity of motion is as mv. After a quick example he continued,

> There appears to be no ground for making a distinction between the *quantity of motion* and the *force* of a body in motion; as all the power or activity of body arises from and depends upon its motion. (p.106)

This use of terms is confusing to the modern reader, since in conjunction with the previous statement Maclaurin appears to be saying $F = mv$. It is worth noting[15] that Newton did not write $F = ma$, but rather, "the change of motion is proportional to the motive force impressed," or at best $F \propto m\Delta v$. If the body started from rest ($v = 0$), then the motive force impressed in the body would in fact be $m\Delta v = m(v_f - 0) = mv$, yielding $F = mv$, leaving aside for the moment how to distinguish between relative and absolute rest.

Hume's apparently erroneous equation of moment and force is found in the text of one of the foremost mathematical Newtonians of the time. Maclaurin's facility with the rest of the Newtonian system allowed him to avoid any difficulties associated with this reading. In particular, Maclaurin later made his discussion of force more precise, and clearly separated proper from improper measures of force, a distinction not followed by Berkeley and Hume.

> 22. The principle, "that the cause is to be measured by its effect," is one of those that will be very apt to lead us into error, both in metaphysics and natural philosophy, if applied in a vague and indistinct manner, without sufficient precautions. Force is defined [by others] to be that power of acting in a body which must be measured by its whole effect till its motion be destroyed, by those who favour the new opinion, or some of of [sic] them at least, and by some who would represent this dispute as merely about words. [75, p.136]

However, Maclaurin argued that Newton had the antidote to such muddy definitions:

> . . . the impressed force being considered as the cause, the change of motion produced by it is the effect that measures the cause; and not the space described by it against the action of an uniform gravity, nor the hollows produced by the body falling into clay. [75, p.136]

[15] There is also a nice discussion of this matter in I.B. Cohen [19].

Unfortunately, while Hume might have copied the former argument about the definition of force from Maclaurin, he omitted the latter about its proper measure.

In section C Hume has "force" where "weight" or "mass" would seem the more reasonable term. As he has it, force is being compared with raising a weight, in this case violating Maclaurin's warning to measure force only with change in motion. Hume appears to have been using ideas more common to Descartes and several post-Cartesian philosophers. For instance, Huygens wrote, "What I mean by forces is the power ... of raising a weight." [39, p.177]. Similarly, Torricelli thought it was possible to equate the "forza" of a 1,000 lb object with that of a 100 lb object dropped from a great enough height. From analyses such as these sprung many debates involving live and dead forces, and paradoxes about the infinities involved in converting one to another. [115]

It is clear what Hume is trying to get at: the faster something moves, the greater its impact. Although Hume uses the phrase, "solid contents," he may mean something closer to "weight." Then in section C, Hume can be read as saying, much more in line with Huygens, "a small force [weight] can raise the greatest weight if we can increase the velocity of the force [small weight]." I believe the correct picture is shown in figure 2.1. In that figure we have a smaller

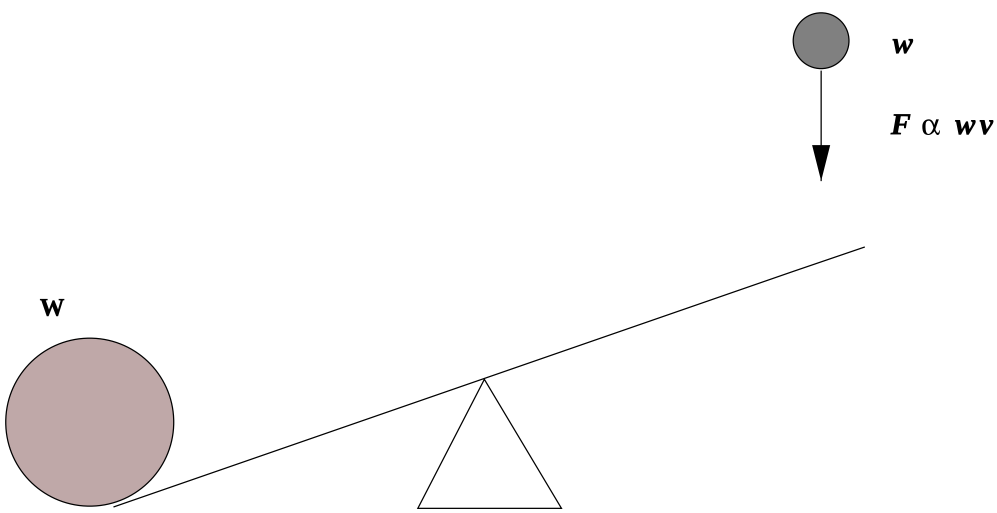


Figure 2.1: Interpretation of Hume's idea of force

weight w dropped onto a lever set up to raise the larger weight W. The total "force" associated with w increases with its downward speed. Hence, to raise a heavier W, using lighter w, one need only increase the speed with which w is projected onto the lever. This will increase the force and the moment, as Hume defined those terms.

If Hume gleaned his Newton largely from Maclaurin, then his mistake is understandable. The best interpreters at the time had not come to a consensus on the exact formulation, and many natural philosophers before Newton had analyzed motion in terms quite similar to Hume's. Moreover, the whole business of dropping weights into clay was part of the ongoing *vis viva* dispute, in which followers of Leibniz argued that the proper measure of force was mv^2. For whatever reason, Hume did not adopt the notion of measuring force by the *change* of motion, nor did he heed Maclaurin's explicit warning against measuring force by the rise of a weight. As it is, Hume's "force" lies somewhere between momentum mv and impulse Fdt.

The *Enquiry* provided a more direct and dismissive reference to Newtonian natural philosophy in Hume's footnote on inertia. [55, p.84,fn.11]

> I need not examine at length the *vis inertiae* which is so much talked of in the new [Newtonian] philosophy, and which is ascribed to matter. We find by experience that a body at rest or in motion continues forever in its present state, till put from it by some new cause; and that a body impelled takes as much motion from the impelling body as it acquires itself. These are facts. When we call this a *vis inertiae*, we only mark these facts, without pretending to have any idea of the inert power, in the same manner as, when we talk of gravity, we mean certain effects without comprehending that active power.

Maclaurin introduced inertia with a lucid evocation of the idea by appeal to common experience, and then used it as a central concept and tool in his development and defense of Newton. He argued that inertia allowed one to distinguish true and absolute motion through absolute space. Were one to stop the earth from spinning, which would require measurable force, the stars would stop moving, though no force had been applied to them; had they really been in motion beforehand, that would violate the law of inertia. Further, the pendulum was known to be retarded at the equator compared to the poles. Both of these phenomena were, "consequences of the real motion of the earth upon its axis." [75, p.102] Similarly, again following Newton, stop a ship and each body on board flies off, "for, in consequence of its *inertia*, it endeavours to persevere, not in its state of rest in the ship, but in its state of motion or rest with regard to absolute space." (p.103) By denying the special status of inertia and absolute space, Hume undermined the Newtonian appeal to the reality of forces and thereby set limits on the nature of causal inference in natural philosophy. Hume had a special interest in cutting of the Newtonian argument at this point: any farther and the system could be used to justify natural theology.

Newton and other followers may have waffled about the status of causes, but Maclaurin was clear on his interpretation. Natural philosophy was about explaining the causes of the phenomena of nature, and most definitely in the service of natural religion. It was important to get things right, because "false schemes of natural philosophy may lead to atheism," as Maclaurin thought it had in Lucretius' "monstrous system." (pp.3-4) Maclaurin argued that through the method of analysis and synthesis, Newton had gone beyond hypotheses to demonstrably true causes, which could be distinguished by appeal to the real forces acting in the world. However, contemplation of universal gravity and the lawful order of the whole system led Maclaurin inevitably to God, and with Newton, to the idea that God was directly behind universal gravity.

With that move, the Newtonians became occasionalists, and Hume boldly accused them of failing to understand Newton.

> It was never the meaning of Sir Isaac Newton to rob second causes of all force or energy, though some of his followers have endeavored to establish that theory upon his authority. On the contrary, that great philosopher had recourse to an ethereal active fluid to explain his universal attraction, though he was so cautious and modest as to allow that it was a mere hypothesis not to be insisted on without more experiments.

Hume may be forgiven for not knowing Newton's unpublished works. Newton's ætherial fluid was at best a half-hearted attempt, and probably put in just to appease mechanists by suggesting that something like an æther might be found to explain gravity. So Hume is right that Newton allowed that even "second causes" of gravity would have power, but contrary to Hume, Newton did put God very close to the nature of gravity.

Force, Cause, God

Maclaurin was a straightforward realist. Our internal consciousness, he said, convinced us "that there are objects, powers, or causes without us, and that act upon us." [75, p.97] Maclaurin saw no problem with the basic representationalist theory of perception, even of cause and effect:

> The mind is intimately conscious of its own activity in reflecting upon its ideas, in examining and ranging them, in forming such as are complex from the more simple, in reasoning from them, and in its elections and determinations. From this, as well as from the influence of external objects upon the mind, and from the course of nature, it easily acquires the ideas of cause and effect. (p.99)

Hume would would have agreed that the mind easily acquires ideas of cause and effect, and that cause and effect are tied to the actions of the mind, but he would show that this process could not be so rational as Maclaurin made it out to be. In addition, Maclaurin's natural theology could not have passed Humean muster. Maclaurin thought the inference to mind, spirit and God were as apparent as those of cause and effect. In a sense Hume agreed: he undermined the ultimate rational justification for *any* inference to cause and effect. However, as already mentioned, *within* the scope of natural philosophy, certain rules of reasoning were allowed.

Hume argued two fronts here, against both Newtonian and Cartesian natural theology: first, that our ignorance of mechanical causes was insufficient warrant for supposing God to be involved in everything; in fact it was no harder to suppose that "motion may arise from impulse" than that God was directly involved. Indeed, the inference to God seemed to Hume an infinitely larger leap than one to a single unexplained impulse. Second and conversely, the fact that we cannot warrant an inference to God did not raise impulse and inertia to the status of real explanations either. Rather, "All we know is our profound ignorance in both cases."

Hume had several arguments against the inference to a Designer. Although he acknowledged that as a matter of fact, the mind was compelled when considering the workings of the universe to imagine a creator, he saw no *justification* for this belief, and thought it had no place in natural philosophy.

- We see but a small part of the universe, and that part poorly. "And do we thence pronounce decisively concerning the origin of the whole?"

- To conclude that an orderly universe "must arise from some thought and art . . . it were requisite, that we had experience of the origin of worlds. . . ," which of course we do not.

- Even granting that like effects require like causes, as we must in natural philosophy, is it really true that the works of nature resemble the works of humanity? Hume thought that the creation of planets and stars was so far beyond that of humanity to be a dissimilarity. Conversely, the closer the resemblance, the closer God must be to human, which was also absurd.

- "Like effects prove like causes," at least within natural philosophy, but ultimately all inferences from experience are tentative and subject to revision: how do we know nature or God will continue to be uniform with respect to causes?

What makes this particularly interesting is that Hume's guiding principles very closely follow Newton's four rules of reasoning.

In the *Treatise*, Hume set forth eight "rules by which to judge cause and effect." He indicated that such rules were especially necessary since by his skeptical arguments, "'tis possible for all objects to become causes or effects to each other," since all connections are equally unknowable. Therefore it was "proper to fix some general rules, by which we may know when they really are so." [53, I.III.XV, p.173] The most relevant rules are the fourth and fifth. Hume wrote, were:

> 4. The same cause always produces the same effect, and the same effect never arises but from the same cause. This principle we derive from experience, and is the source of most of our philosophical reasonings. For when by any clear experiment we have discover'd the causes or effects of any phænomenon, we immediately extend our observation to every phænomenon of the same kind, without waiting for that constant repetition, from which the first idea of this relation is deriv'd.
>
> 5. There is another principle, which hangs upon this, *viz.* that where several different objects produce the same effect, it must be by means of some quality, which we discover to be common amongst them. For as like effects imply like causes, we must always ascribe the causation to the circumstance, wherein we discover the resemblance.

Compare, for instance, Newton's second rule of reasoning in philosophy:

> *Therefore to the same natural effects we must, as far as possible, assign the same causes.*

In Hume's *Dialogues Concerning Natural Religion*, Hume's proponent argued against the application of this rule for theological ends. [56, Pt.5], although, significantly, without attacking it directly. First he questioned the resemblance of the effects in the design argument, claiming that the universe did not really resemble human artifacts. Then, perhaps using a variant of Rule IV, Hume suggested that Rule II was invalidly applied in this case. Recall that according to Newton's Rule IV, any propositions of experimental philosophy deemed "accurately or very nearly true," may have to be revised at "such time as other phenomena occur" which contradict the propositions. As Newton wrote it, Rule IV stated,

> In experimental philosophy we are to look upon propositions inferred by general induction from phenomena as accurately or very nearly true, notwithstanding any contrary hypotheses that may be imagined, till such time as other phenomena occur, by which they may either be made more accurate, or liable to exceptions.

Force pointed out that this is a principle of the fallibility of induction from experience. [34] Hume's attack cleverly used the Rule in reverse. Since the design argument could not be so tested by experience, Hume concluded that it did not belong to experimental philosophy, contrary to what Newtonians wanted to claim. To add insult to injury, Hume made it clear that by defending the design principle, Newtonian philosophers were following the Cartesian model of hypothesizing, which of course Newtonians wanted to reject. By inferring without a doubt the existence of a designer, Newton was, according to Hume, guilty of going beyond his own modest rules of reasoning. It was Newton's adherence to these rules elsewhere in his natural philosophy that made Hume esteem Newton above all others, and to which Hume attributed Newton's great success.

Hume did not disagree with Newton's second rule, but only its application in areas not amenable to the experimental philosophy.

> If we carry our enquiry beyond the appearances of objects to the senses, I am afraid, that most of our conclusions will be full of scepticism and uncertainty. Thus if it be ask'd, whether or not the invisible and intangible distance be always full of body, or of something that by an improvement of our organs might become visible or tangible, I must acknowledge, that I find no very decisive arguments on either side; tho' I am inclin'd to the contrary opinion, as being more suitable to vulgar and popular notions. If *the Newtonian* philosophy be rightly understood, it will be found to mean no more. A vacuum is asserted: That is, bodies are said to be plac'd after such a manner, is to receive bodies betwixt them, without impulsion or penetration. The real nature of this position of bodies is unknown. We are only acquainted with its effects on the senses, and its power of receiving body. Nothing is more suitable to that philosophy, than a modest scepticism to a certain degree, and a fair confession of ignorance in subjects, that exceed all human capacity. [53, Appendix, pp.638-9], noted in [34, p.170]

By applying Newton's methodology to aspects of Newton's system itself, Hume undermined some of the tenets of the whole metaphysical package held by Newton and his followers, as exemplified by Maclaurin. Even though "same cause same effect" was a judicious (although unjustifiable) rule for reasoning in natural philosophy, its scope did not extend beyond the proper domain of that science, and could not license metaphysical inferences about God, nor even inferences from a universal generalization to a natural law.

Thus robbed of the special import of "inertia," "force," and "absolute reference frame," Maclaurin and Newton's defenses of the real superiority of their system over merely relative ones was less convincing. We know from his letters to Bentley [108, pp.46-58] that Newton himself was not entirely happy with the implications of his philosophy, but he remained committed to his deviations from the mechanical hypotheses, and as we have seen, even added his fourth rule of reasoning to defend those deviations.

Summary of Newton & Hume

Hume's philosophy clearly derives from Newton's natural philosophy and is both an attempt to extend the Newtonian method into the moral sciences, and an attempt to set the proper bounds of reason in order to preserve the progress of science without letting it mix with theological metaphysics. The second effort sought to cut through the metaphysical fog which had surrounded the Newtonian philosophy since its inception, by delineating clearly what could and could not be known, and how so. Like Locke, Hume wanted to shine a light in the philosophical darkness and clear away the speculative rubbish which appeared to have hindered the moral sciences, and left them far behind natural philosophy. To this end he championed the Newtonian reservation against pursuing ultimate explanations, and the preference for empirical knowledge, even if its epistemological warrant were less secure than that demanded by Aristotle and Descartes.

Like Berkeley, however, he wanted to curtail the metaphysical excesses he saw present even in the new philosophy, and to ensure scientific progress by demonstrating what questions could not be fruitfully pursued within the scope of an empirical or experimental philosophy. His foremost target in this respect was the natural theology licensed by Newton's success and underscored by that scientist's great authority. Secondly, he wished to limit the speculation on the nature of force and gravity, taking Newton's defensive "it is enough" as a prescriptive ward against proceeding farther than the first level of descriptive causation. Together with his incisive and general critique of causality and induction, Hume helped to define the goals and limits of modern science.

Hume's brand of argument has become an integral part of modern science for exorcizing more than religion. Indeed, the behaviorists used similar arguments to remove talk of "mind" from psychology, and the Churchlands, not themselves behaviorists, still do. However, the Humean purification of science has left behind a philosophical void. Hume's skeptical arguments against metaphysical inferences also barred the possibility of causal inferences. Nevertheless those inferences happened all the time. Hume's positive account, to which we will now turn, was not up to the task of replacing what had been lost. To the philosopher and psychologist fell the problems of formulating a replacement for Hume's constructive account, and figuring out just what the mental process of causal attribution was. We will first look at replies to Hume in recent philosophy of science, and then turn to psychological investigations of causation and event perception, which we shall see also begin with Hume.

Chapter 3

Hume's Ghost: Modern Philosophy & Causation

Hume's ghost continues to haunt the modern study of causation. First, it urges the search for empirical accounts of causation. Second, it makes them impossible to find. According to Hume, for something truly to deserve the name "cause," it had to involve a necessary connection. But Hume also held that ideas came from sense impressions, and there can neither be a sense impression for "necessary connection" or for "causality," so the idea of causation was relegated to the nether regions of psychology. However, as the next chapter will examine in detail, perception encompasses more than the positivist sensation theory Hume accepted. If perception can extend beyond sensation, even to events, then causation may at least sometimes be observable. The next chapter will turn to the psychology of event perception. Before getting to that, some philosophical work remains.

Regardless whether events are perceivable, necessary connections are not. In order to make room for the possibility of causal perception, we must relinquish the idea of necessary connection. A survey of the debate will show additional features to be removed. Chief among them is what Paul Humphreys has labeled "passive empiricism," a philosophical malady which Hume has transmitted to many of his readers. This is the empiricism of an observer who cannot in any way intervene in or interact with the world, but is merely acted upon and must make inferences based on nonparticipatory correlations. It includes an excessively weak idea of experiment, which for Hume was barely distinguishable from "experience," a poor and nonquantitative description of events, and an avoidance of empirical work. In the "Rules of Reasoning," where Hume made some progress towards a methodology of causal attribution, he himself abandoned some of the classical Humean themes.

Hume understood that causation was "the cement of the universe," and despite his general skepticism, allowed that in natural philosophy it was natural and necessary to adopt the rule, "same cause, same effect," even though no ultimate justification existed for so doing. [22] His rules of reasoning assumed that one could take this step, and even seemed to allow for the possibility of finding a cause in a single experiment. However, as Ducasse pointed out, that possibility requires a completely different account of causality and the possibility of causal perception than Hume's "constant conjunction."

Causation is also central to science in general. This fact is not as apparent when one looks at experimental physics, where control is strict and cause and effect are rarely a question. However, as one moves towards observational disciplines like field biology, environmental science, psychology, and economics, causation becomes a methodological concern. Why? Precisely because for observational science, it is so hard to pin down the causes and separate them from mere correlations and other artifacts. Scientists working in biology and econometrics have developed tools to help them trace genuine causal connections, and some of these tools have in turn proved very helpful for philosophers. [59]

There are two ways, then, that philosophers can benefit from examining real, empirical, causal inference. First, they should track actual scientific practice. The major advances in twentieth-century philosophy of causation have come from the probabilistic theorists who have increasingly incorporated the methods and ideas of causal modeling, and the process theorists who have reformulated the idea of causation in a spacetime ontology and looked for real events which stand in causal roles.

The second way to benefit from empirical work is to study ordinary causal perception. In other words, pursue the goal of Hume's *Treatise*, to bring the experimental method of reasoning into philosophy. There are limits to this path, for any perceptual system is subject to illusions, and perceptual causation should not be expected to be equivalent to objective causation. However, since the role of causation in science is to provide explanations and increase our understanding, the causes we find and deem useful will be structured according to our own perceptual understanding of causation, which can best be studied empirically. If some instances of causation can be perceived, and these form the basis for our ideas of causation, then there is an answer to, or at least a dissolution of, Hume's central problem. A collateral benefit to the empirical approach to actual causes is that by eliminating the need to provide some account of causality which works in all possible worlds, we eliminate a lot of baggage and hopefully reduce the amount of time we spend conjuring up and evaluating toy examples.

This chapter begins with a quick survey of the current debate in the philosophy of causation, and then develops in detail the singularist, process-oriented approach of the conserved-quantity (CQ) theory. That account provides the philosophical grounding for the psychological work in event perception presented in the following chapters, which in turn provide evidence for the possibility for causal perception. I have postponed in-depth discussion of probabilistic approaches to causality, including causal modeling, for the last chapter, where they are presented alongside relatively recent work in cognitive psychology which has begun to adopt and explore that approach.

3.1 Problems in Hume

By removing the single event from consideration, Hume had available at most constant conjunction, priority and contiguity. According to Hume, constant conjunction plus priority and perhaps contiguity was sufficient for the impression of causation or necessity. A constant conjunction is a universal statement, "All X's have been Y's." In the limit, "All X's are Y's." This universal generalization is a lawlike relation, because its expression is in the form of a lawlike statement. Because his account left no other basis for making distinctions, Hume asserted that all lawlike relations are the source of causal impressions. The problem is that this is too inclusive. Night always follows day, bachelors are always unmarried, and all known or suspected samples of diamond have a mass of less than 500 kg. Yet none of these constant conjunctions of properties or events is a causal relation. In addition, the account cannot distinguish two common effects of a single cause ($A \leftarrow C \rightarrow B$) from two events which are cause and effect ($A \rightarrow B$).[1] Lawlikeness is insufficient.

It might be possible, however, that some *subset* of lawlike relations would be restrictive enough to pick out only the causal relations. If we could find that subset, then we could augment Hume's own account and keep his commitment to lawlike causes. Alternatively, we could defend the possibility of singular causes and avoid the questions of laws. Either way, we must recognize that there are two distinctions to be drawn. The first is the distinction between lawlike versus accidental relations. The second is between causal and noncausal relations. Hume's theory was too inclusive because it identified the two distinctions. In fact there are four possibilities:

	Lawlike	**Singular**
Causal	1	2
Noncausal	3	4

Assuming that all boxes are occupied, the goal is to find a way to distinguish items in boxes 1 and 2 from those in boxes 3 and 4. The exclusion of items from box 4 is easy, and Hume's constant-conjunction theory did the job. However, it incorrectly included members of box 3. Of course it also excluded members of box 2, but Hume did not believe that any existed. For clarity, here is the grid with candidates for memberhood filled in:

	Lawlike	**Singular**
Causal	Regular causation	One-time causal interactions
Noncausal	Accidental generalizations	One-time coincidences

So, Hume's account was too broad because it included accidental generalizations. According to some philosophers and psychologists, Hume's account is also too *narrow*, in that it excludes singular causes or singular causal perceptions. Hume suspected this problem himself when he noted that some learning takes place without repetition of any relevant sort. His example was of a child who burns himself in a candle. Such a painful experience usually need not be repeated before the inference is made, and for Hume that kind of inference is the essence of causality. As we shall see in the next chapter, psychologists beginning with Michotte and Johansson pursued this line to find out what visual properties in the world are actually available to us and from which we can draw causal inferences. Neo-Humeans may argue that these items do not truly belong in box 2, because to the extent they are causal relations, they must be lawlike, even if manifested only a single time. That move relinquishes a truly empiricist idea of laws, but it is possible. Regardless of the status of box 2, some work remains if we are to exclude members of box 3.

The possibility of probabilistic or indeterministic causes comprises a completely separate challenge to the Humean reliance on necessary connection. In all fairness, Hume suspected that conjunction need not be constant in order to support a causal impression, but merely "for the most part." He did not develop this, perhaps because he was not inclined towards mathematical analyses, but he did not ignore the possibility. He presented the option, and then told the reader that he would nevertheless proceed to treat only the basic case. Hume's statistical concession is worth mentioning, but it is not a very deep concession. Although humans are tolerant of some degree of *non*constant conjunction, Hume thought it the mark of a *philosopher* to believe in the constancy of causes. For

[1] Throughout the text I shall use arrows ($\rightarrow$) exclusively for causal relations.

the philosopher, the observed nonconstancy was to be explained by unnoticed counteracting causes. Such a move already identified causation and necessary connection, and twentieth-century philosophers have denied that claim. This thesis will examine indeterministic causality as it is fit into process theories and the direction of causation, but omit a deeper treatment of the probabilistic framework.

3.2 Alternatives to Hume

It is universally granted that Hume's skeptical attack against the rationalists was successful. Having shown therein that causes could not logically necessitate their effects, two avenues were open to Hume:

1. Decide that something was wrong with the "necessary connection" definition of cause, since as it stood, it could not possibly be satisfied.

2. Keep the definition and deny the reality of causes.

Hume chose the second option.

There are episodes in history were a brilliant thinker points towards an insightful reformulation of a classic problem, and then walks away from that insight. Hume's move at this point is like Parmenides' decision to deny the reality of motion because his theory said it was impossible. Having chosen the second option, Hume had to give a reductionist account to explain causal impressions, which now no longer corresponded to anything real in the world. Most modern philosophers followed this same program. Broadly categorized, modern accounts of causation fall into six groups[2]:

1. Regularity & Nomological accounts

2. Necessary & Sufficient Conditions accounts

3. Counterfactual accounts

4. Process accounts

5. Probabilistic accounts

6. Manipulation accounts

The list is roughly ordered by fidelity to Hume's theory. The first three better-known options stay closest to Hume's own program. The last three depart from Hume in fairly significant ways. I will start with thumbnail sketches.

3.2.1 Humelike Alternatives to Hume

Regularity & Nomological

Hume's "constant conjunction" account is an example of a "raw" regularity theory: causation is reduced entirely to regularities. Cartwright has described this part of Hume's approach as a "two-step" program: first ground singular causal facts in generic causal facts, second reduce generic causal facts to regularities. [13] As already mentioned, the obvious problem with Hume's theory is the inability to separate accidental generalization from natural laws. The obvious

[2] Some philosophers have also analyzed the logic of the causal relation without giving an account of causality, but that is an entirely separate project.

solution is to supplement Hume's theory with natural laws. Such solutions are called nomological accounts.

Nomological accounts take as given the existence of natural laws, and reduce causation to deducibility from the totality of noncausal facts plus natural laws. By fiat, this solves the problem of accidental generalizations, which is a major advance. The immediate problem is that not all laws do the trick. The standard counterexample is that the ideal gas law, $PV = nRT$, is not causal. All it describes is functional covariation, or as Nagel [83] put it, concurrent dependence. It violates the intuitions that cause and effect are asymmetric, and that causes should precede their effects.

One reply is to narrow the class of law which can be used. For instance, perhaps dynamical laws, which Nagel called "laws of temporal dependence," would be more appropriate? For instance, take one of the laws of motion, $\ddot{x} = gt$. This kind of law specifies the state at some future time given the current state, and appears to solve our problem. However, the laws of motion are notorious for being reversible. [83] In fact, astronomers run them backwards all the time to retrodict the path of comets, or to examine the origins of the solar system or the universe. It may be that causal laws are dynamical, but dynamical laws are not necessarily causal.

There are two ways to go from here. First, nomological theorists could accept the result of their theory and claim that causes are symmetric and possibly simultaneous. I think this is to make the same mistake as Hume and Parmenides. At any rate, it would represent a radical departure from Hume. The other option is to look for further refinements that will distinguish *causal* laws from noncausal laws. Unfortunately, this is the original problem, only now at the level causal laws instead of causal relations. As Nancy Cartwright has pointed out, if the goal of this reduction is to satisfy our empiricist cravings, it would probably be more straightforward to cut straight to the reality of causal relations rather than posit the reality of some sort of causal laws.

Necessary & Sufficient Conditions

The conditions approach traces its pedigree back to Hume's "rules of reasoning in philosophy," and Mill's improvements thereupon. Causes are held to be necessary conditions for their effects, sufficient conditions for their effects, necessary *and* sufficient conditions, or some other combination of these conditions. The most successful version is Mackie's INUS account: a cause may be in itself sufficient or necessary, but it is *at least* an INUS condition, an Insufficient but Necessary part of a condition which is itself Unnecessary but Sufficient. Mackie's account captures the methods of diagnostic troubleshooting very well, but it inherits some shortcomings of the necessity and sufficiency approaches.

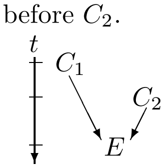
First, a cause cannot be equal to a necessary condition. My laptop failed to turn on (call this effect L), because it was unplugged (call this condition U.) However, there were a number of other ways to cause L, so U was not a necessary condition for L. Second, neither was U sufficient for L: my laptop can run off its battery. So in addition to U, the battery had to be discharged, or absent, or malfunctioning in some way. Call that condition B. Was the joint condition $U + B$ necessary? No, there are any number of ways a computer can fail. Sufficient? No. At minimum you have to specify that my laptop had no other conceivable way to get power. And if we reverse the example and try to explain why it turned on ($\bar{F}$), it becomes clear that much more than $\bar{U} + \bar{B}$ is required for sufficiency.

What about INUS conditions? Was U an INUS cause for L? Yes. It was an insufficient (already proven) but necessary *part* of a total condition (the state of the battery and everything else) which was itself unnecessary (other configurations including power outages could cause L) but sufficient (given the actual state U and everything else relevant) condition. Mackie's INUS account is a very good aid for *finding* causes. However, it has a couple of problems as a *definition* of a cause.

If a cause is necessary and sufficient for its effect, then the effect is necessary and sufficient for its cause. Since causation is supposed to be asymmetric, standard regularity accounts fail. The bare INUS theory is not perfectly symmetric, but it is close.

> It is logically possible that A should be an inus condition of B while B is not an inus condition of A. But ... in fact where A is an inus condition of B, B is usually also an inus condition of A. [74, pp.160-1]

For several reasons, Mackie chose not to rely on the direction of time to give the direction of causation. Instead, he added a condition that the cause must be "causally prior" to the effect. Causal priority was understood in terms of "fixity": C is fixed prior to E if there cannot have been a time when E was causally fixed but C was not. Clearly this requires a measure of indeterminism, but so will our solutions, so that is no objection. Unfortunately, fixity fails to resolve causal asymmetry at least in the case of overdetermined effects. In the following diagram, the overdetermined effect E was fixed back when its first cause C_1 happened, which was before C_2.



However, that means that E was fixed before C_2. By supposition, C_2 is an INUS condition for E, and therefore it is very likely that E is also an INUS condition for C_2. Since they are both INUS conditions for each other, the one fixed first is the cause, but in this case that one is E. Therefore by supposition, $C_2 \rightarrow E$ but by derivation, $E \rightarrow C_2$, and causation becomes symmetric. [10]

The nature of indeterminism in the fixity account is ambiguous. There has to be some, else all events are fixed and there is no causal direction. However, an indeterministic cause as such can never be an INUS condition, at least until its outcome is determined: if it were a necessary component of the actual condition, then by the indeterminism, that condition would not be sufficient. Paul Humphreys has several very convincing examples of the pernicious pervasiveness of indeterminism in macroscopic systems, which make this objection quite strong. [58, pp.14-16]

As Mackie himself has illustrated, the bare INUS account cannot distinguish two common effects of a single cause ($A \leftarrow C \rightarrow B$) from cause and effect ($A \leftrightarrow B$). Fixity alone does not help, but it could be used to rule out such cases. If synchronous fixity among events disqualified any direct causal links among those events, then in the case of ($A \leftarrow C \rightarrow B$), ($A \not\leftarrow C \rightarrow B$) because there is never a time when A is fixed but not B, and conversely. Unfortunately, this still does not solve Mackie's famous counterexample.

In that example, factory workers in Manchester and London leave work after hearing the end-of-work whistle (hooter) go off. However, Manchester and London are at some distance. Surely the blasting of the Manchester whistles is not the cause for the L, the Londoners' leaving work. Unfortunately, given the fundamental role of regularity in the INUS account, the sounding of the

Manchester whistles is an INUS condition for L, and of course it is fixed prior to L. The problem, as Kim pointed out, is that the INUS account reduces to other variants.

> Singular causal assertions are explained in terms of the [INUS condition]. The notion of INUS condition in turn is explained on the basis of 'necessary condition' and 'sufficient condition', and these are analysed in terms of counterfactual conditionals. Finally, counterfactuals are explained on the nomic-inferential model. It is at this point that laws and regularities enter into singular causal judgements; according to Mackie, his analysis can be characterized as a form of the regularity theory of causation. [63, pp.62-63]

Therefore, ultimately, Mackie's account must also deal with any remaining objections to regularity theories not specifically addressed by the INUS formulation.

Counterfactual

At one level, counterfactual accounts are clear and simple: E causally depends on C just in the case that E would not have occurred if F had not occurred. Then, C is a cause of E iff there is a chain of events from C to E such that each event in that chain causally depends on its predecessor. Counterfactuals seem an obvious solution to the problem of distinguishing accidental generalizations from natural laws. Natural laws support or allow counterfactual claims, such as,

> If these two barely subcritical masses of plutonium had been brought together they would have undergone fission.

Accidental generalizations do not allow such counterfactual claims. Even if all the items in my pocket are, always have been, and always will be metallic, we could not say,

> If this handkerchief were put in my pocket, it would become metallic.

However, it is not always obvious which counterfactual a natural law supports. There are natural laws at work assuring that any naturally-occurring samples of plutonium we find will be subcritical. But we should not infer,

> If these two barely subcritical masses of naturally-occurring plutonium had occurred together they would each have been smaller.

Instead of using counterfactuals to try to sort out natural laws, one could take counterfactuals to be basic, using the definitions given previously. However, counterfactual dependency, so defined, is much broader than causal dependency. For instance, analytical statements can be counterfactually dependent: a bachelor who gets married is no longer a bachelor, but that was not a causal change. Second, the account cannot handle overdetermined events, since in that case none of the causes are counterfactually causes, so any overdetermined event ends up being completely uncaused. Lewis, the most prominent proponent of pure counterfactuals, accepted this problem and replied that overdetermination never happens. There is something to this objection. Computer technicians do not have to evaluate all possible causes for a computer that is not working properly, because they are rarely diagnosing some abstract state called "not working properly". Different problems result in *different* failures, and compound problems in still different or compound failures. So some cases of supposed overdetermination will disappear with proper redescription, and perhaps they all will.

Even if these problems can be surmounted, the most fundamental objection to counterfactual accounts is that there is no useful way to evaluate their truth or falsity. The best existing account of counterfactuals [72] suggests comparing possible worlds, but there is no agreed upon metric for closeness of possible worlds, let alone agreement about how to quantify properties and their differences. Last, the very counterfactuals used to describe causation, "the effect would not have occurred if the cause had not occurred", are in fact statements of necessary conditions, which we have already reviewed. The counterfactual account does not help unless there is a convincing account of counterfactual semantics.

Korb has concisely diagnosed the common failing of all the alternatives presented so far.

> Since all of these analyses assume at core that there is nothing more to causality than universality, all of them have failed in the end to exclude accidental or spurious relationships. As a symptom of such failure, they fail in some cases to distinguish between causes and effects — that is, what is obviously an *effect* may well satisfy the conditions these analyses put upon causes. [67, p.2]

In other words, these accounts have all followed Hume down the wrong fork in the path: they assume that causation is some form of necessary connection. The remaining options take the other path with varying degrees of success. That is, they look for causes as something *in the world* which corresponds to causation. Some trace that something to some sort of directly observable property. Others, in particular propensity theories, just assert the existence of objective chances and proceed from there.

3.2.2 UnHumelike Alternatives

> There is more to causality than regularity; the causal connections furnished by causal processes explain the causal regularities we find in the world.
>
> Salmon

Of the three unHumelike accounts of causality, this thesis only really addresses the process theory. While developing the process theory, we will have to introduce probabilistic or indeterministic causality, but a full treatment of probabilism will not be needed to make sense of the direct perception thesis. Indeed, whatever they are, propensities and probabilities are not directly observable!

Instead, in this chapter I will concentrate on exploring the process theory from its basic singularist insight to the modern CQ version. I will not give a historical reconstruction, which would require a detour through statistical relevance, mark theory, and more counterfactuals. The next two chapters will rely on a physical theory of causality, of which the CQ theory is the best option. However, it currently exists in two slightly different flavors, and it will be worth going through those in some detail.

I will skip the manipulability account largely because it is only a clarification of the idea, not a theory of causality. This account claims that C was a cause of E if a person manipulated (something like) C to make (something like) E occur. In the subjunctive form, what is required that C could have been brought about in order to bring about E. No doubt that is true, and it perhaps serves as an uncontroversial base case, but it does not clarify the nature of the connection. It does not allow us to distinguish between accidental and causal correlation, and it uses causal notions in the definition of causation. Aside from the appeal of throwing out the whole problem by asserting that causality is an anthropomorphic idea, rather than a feature of the world, manipulability theory is not an account.

3.3 The Singularist Critique

According to Ducasse, Hume's fundamental epistemological argument can be broken down into three parts. [29]

1. To be is to be perceived.
2. No connection between cause and effect is ever perceived.
3. Therefore there is no such connection (but only an internal impression of the mind).

Because of his own experiences, Ducasse chose to deny the second premise.

> I bring into the room and place on the desk a paper-covered parcel tied with string in the ordinary way, and ask the students to observe closely what occurs. Then, proceeding slowly so that observation may be easy, I put my hand on the parcel. The end of the parcel the students face then at once glows. I then ask them what caused it to glow at that moment, and they naturally answer that the glowing was caused by what I did to the parcel immediately before. [29, p.71]

When Ducasse asked his students to justify why they thought that what he did caused the glow, Ducasse's students said that "nothing else happened to the parcel at the time."

Apparently, from a single case, observers determined the link between cause and effect, and gave a criterion for making the distinction. In the next chapter we will see more evidence for singular perception. If these cases do constitute counterexamples for Hume, how do they do so?

As it turns out, merely denying the second part of Hume's argument is insufficient, because it is tied to the first part. And, to be clear, Hume did not make the Berkeleyan claim that "to be is to be perceived." However, he did claim that to be *known* is to be perceived. All ideas came, ultimately, from sense experience.[3] This means that for Hume, we could not have the idea of causality unless it were perceived, and there were two obstacles to its perception. First, "perception" for Hume was "sensation," and sensation was reserved for very primitive sorts of data, like his "coloured points:" minimum sensibilia of some definite extent and quality. In other words, for the causal connection to be perceived, on Hume's account of perception, there would have to be a sensation of this *thing* called a causal connection, *separate* from the events or objects being connected. Second, if that were not difficult enough, in order to count, whatever it was would have to be a sensation of a *necessary* connection, which of course was impossible by virtue of the fact that empirical knowledge could never reach certainty, a fact about which Hume unfortunately had to remind philosophers.

Whatever we make of Ducasse's chosen example, I want to claim that he had pinned down the central problems in Hume's positive account, or lack thereof,

[3] Except some ideas like space and time and related mathematical concepts. See [35].

of causality. What is even more interesting, however, is that Ducasse noticed that Hume himself admitted as much in a few scattered references. First, he acknowledged that

> not only in philosophy, but even in common life, we may attain the knowledge of a particular cause merely by one experiment, provided it be made with judgment, and after a careful removal of all foreign and superfluous circumstances. [53, p.104]

Merely one experiment! Now that Hume was making rules for reasoning within natural philosophy, he apparently allowed for some inferences to happen in the absence of constant conjunction. He then resolved the incipient contradiction by denying that single experiments were what they claimed to be:

> tho' we are here suppos'd to have had only one experiment of a particular effect, yet we have many millions to convince us of this principle; *that like objects, plac'd in like circumstances, will always produce like effects*.... [53, p.105]

Hume's motivation is sound. Even six-month-old infants have a wealth of experiences with the world, and in the absence of an account of similarity and categorization, we cannot say whether the apparently novel example really does fit under some abstracted pattern of expected motion. However Ducasse observed that Hume's italicized principle did not distinguish causal sequences from accidental sequences. Although it was invoked to explain why causal inferences can in some circumstances happen after an apparent single experience, it would explain equally well any single illicit inferences based on nonlawful generalizations.

In addition, Hume sneaked rather a lot in through the partly empiricist, "removal of all foreign and superfluous circumstances." But this meant, remove those circumstances which are not the cause, and for Hume that in turn meant those which were not constantly followed by the effect. Ducasse observed,

> Preliminary removal of the circumstances which are 'foreign and superfluous' therefore amounts to a preliminary discovery of the circumstance which *is* the cause! Thus, the principle is good not for discovering the cause in a single experiment, but *only for generalizing it* if we have already managed somehow to discover it by a single experiment. If, however, causation can be ascertained by a single experiment, then causation does not consist in constancy of conjunction even if it entails such constancy. [29, p.73]

Ducasse has made the case that at least sometimes it is possible to observe the single cause. At the very least, this should be possible if there was only one noticeable change in the region immediately prior to the effect, and that change was the cause. In the unrealizable ideal case where there really wsa only one change, then Ducasse's theory would be that the particular change C caused the particular change E just when C "alone occurred in the immediate environment of [E] immediately before." [30, p.127] Since there really was only one change, then that change was the cause. As long as it was also perceived as the only change, it will be judged as the only cause. In Ducasse's classroom demonstration, the causal connection was clear, even though the manner of the connection was not understood. On the other hand, the account generalizes poorly.

In reality, there is never only one change. That is, to say the least, unhelpful. On the ontological level Ducasse maintained his theory only at the expense of identifying the cause with the *total* spatiotemporal volume preceding the effect, limited only by the inclusion of a reference class or environment without which the causal question is incomplete. If one makes a cut in the time axis through some region of space, the whole preceding volume[4] is the cause of the whole succeeding volume. In Ducasse's own example of the breaking window, he must accept the sound of the canary song received at the window as part of the cause just as much as the brick which went through it. Ducasse may be complimented for his consistent and successful treatment of "sufficient to." However, it is unhelpful. In the end, it identifies "before" as the cause of "after."

It faces similar problems in epistemology. Ducasse's specific definition was an attempt to replace the use of the word "cause" in everyday use, but it does not do so. When a brick crashes through the window, on the one hand,

> When any philosophically pure-minded person [someone who has not been contaminated by Hume] sees a brick strike a window and the window break, he judges that the impact of the brick was the cause of the breaking, *because* he believes that impact to have been the only change which took place then in the immediate environment of the window. [30, p.131]

On the other hand, Ducasse admitted,

> we cannot without a contradiction refuse to take into the cause *any part* of the total change observed in the contiguous space-time environment of the effect; while, on the contrary, we very frequently in fact seem so to use the word 'cause' as to do just that.[30, p.133]

Return to the first hand, it is not true that observers must believe a cause to be the only change. Observers who could hear the canary song also believe there are sound waves hitting the window. They may even *see* a katydid walking on the window. Nevertheless, they will still judge that the impact of the brick was the cause of the break. This is supported by psychology experiments where observers asked to judge causes discounted more highly correlated changes in favor of more mechanical changes. [2].

Conversely, the account falsely predicts that observers must take C as a cause if it was the *only* change they perceived just prior to E. Now, people are subject to causal illusions, and this account handles them well, but it is incomplete. Observers can judge the only observed change to be irrelevant, and profess ignorance of the cause. If we believe Michotte's experiments, detailed in the next chapter, then a single change in an object A was not always regarded

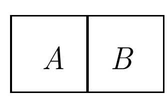


Figure 3.1: Michotte experiment

as the single change in object B. For example, if A suddenly changed color from red to green just before B began to move, the situation was not judged to be causal. Likewise, if A impinged on B from the left, and B departed along a perpendicular path. [80, pp.101-102, 234] Just as some changes are actually more relevant, some are perceptually more relevant.

[4] Or as Michael Friedman has pointed out, the backwards light cone.

Ducasse thought he could avoid the problem by distinguishing the question of, "what caused the effect then and there," which was always the spatiotemporal totality, from "what part was minimally sufficient?" But how do observers know what to ignore so as to leave the minimally sufficient condition? They appeal to "*what was common to the individual causes* of two or more individual events of a given sort." Since this returns us to Hume's regularity account, we can regard this as a failure.

However, Ducasse clarified several issues. First, he emphasized that cause is a relation between events, and that events were themselves extended in time and space. Because causes are relations, and not sensations, Hume was mistaken in trying to build a constructive account on some kind of sensation of causation. For Ducasse, no additional "thing" like force or power is perceived. Rather, one perceives the changing relations between or among events. How that property, causation, is perceived is an open question, just like the question of how relative depth is perceived. It is an empirical question. Since it is about relations between events, the natural field for the study of the epistemology of causation is event perception.

Nevertheless, it would be helpful to have some guide about what properties to investigate. A return to ontology will help here. What is wrong with including the canary song as part of the cause? It is *not* just that canary songs are never found by themselves to break windows, as the regularity solution would have it. That is merely evidence. The reason a song or a nerf ball cannot break a window is that neither has enough kinetic energy to break the bonds keeping the window in the form of a glass. Ducasse opened the door by considering physical processes, but did not take the next step. One way to distinguish the relevant from the irrelevant changes is to incorporate some of the ideas from the physical reductionist theories of causation. Once again, this will help in the goal of being able to compare causes. Let a window be broken by a handful of thrown pebbles, the total sum of which is necessary for the break. On Ducasse's sufficiency account, no proper subset of the pebbles would be a cause. On the accounts we turn to, each pebble would be a contributing cause.

What we take from Ducasse is the possibility of singular causal perception (although perhaps conditional on experience or conservation laws), an event ontology, and an insistence on using extended events, thereby at last discarding the Humean cinema scope. As we shall see, if we are committed to physical processes and durations, then we are squarely in the realm of dynamics, and dynamic properties become crucial in determining and distinguishing actual causes from mere correlates or accidentals.

3.4 Causation and Conserved Quantities

From our point of view, Ducasse's singularist account of causation had two major flaws. The first was that a cause was an entire (possibly very large) spatiotemporal volume and that one could get more specific causes only by appealing to regularity. That left us back at Hume's original problem. The second problem was that for causal perception or causal attribution, observers had to notice or believe that there was only one change prior to the effect. It would be very helpful if there were some way to rule out from consideration canary songs and other mostly irrelevant changes. However, the strict empiricist can do this only through regularity theory and other forms of logical analysis, which we have seen are mostly unsatisfactory.

In 1979, David Fair wrote an article which suggested a clear path out of these analytical tangles: the "physicalistic reduction of the causal relation to one of energy-momentum transference in the technical sense of physics."[31] About a decade later, Phil Dowe [24] used a version of Fair's account to patch well-known holes in Wesley Salmon's process theory of causality [97]. In a series of articles, [24, 98, 25, 99], Dowe and Salmon jointly worked out two closely-related versions of the Dowe's CQ theory, which will both be examined below.

3.4.1 The Account

The CQ theory does not consider "causes" per se, but divides causation into causal *interactions* and causal *processes*. The main purpose of the CQ theory is to distinguish between genuine causal processes like baseballs and pseudo processes like shadows. Salmon thinks that in order to met this goal it is necessary to introduce a third concept, that of *transmission*. Dowe rejects this addition because in his view the work done by transmission is already taken care of given the notion of "object" used in his theory. No event can be the cause of another event unless the two events are linked by a series of processes or interactions. The account does not assign priority to any particular cause. Both versions consider interactions to be logically prior to processes. We will discuss the application of the theory after working through points of agreement and disagreement between the two definitions.

Dowe's Definitions

I will start with the most recent version of Dowe's definitions[25], which have changed slightly through his interaction with Salmon.

Definition 1 (D1) *A* causal interaction *is an intersection of world lines which involves exchange of a conserved quantity.*

Definition 2 (D2) *A* causal process *is a world line of an object which possesses a conserved quantity.*

There are a couple of supporting definitions which deserve mention.

Definition 3 (CQ) *A* conserved quantity *is any quantity which is universally conserved. Current scientific theory is our best guide as to what these are. Good candidates are: mass-energy, linear momentum, angular momentum, and charge.*

Definition 4 (Obj1) *An* object *is "any old gerrymandered thing" that is "wholly present at a time".*

This is equivalent to:

Definition 5 (Obj2) *An* object *is any space-wise but not time-wise gerrymandered thing.*

Definition 6 (WL) *A* world line *is the collection of points on a spacetime diagram which represents the history of an object.*

First, "world line" is clearly just shorthand for "world worm" or "4-volume." This clarification implies that many intersections of world lines such as those representing colliding objects are *partial* intersections. Intersections of waves are not interactions because no exchange is involved, and consequently they may exhibit total intersections. Likewise, pseudo processes like shadows may intersect fully. However, since absorption and emission phenomena also involve complete intersections, the shape of intersection will not suffice for demarcating mere intersections from interactions.

Second, it is not clear whether D2 means that the world line or the object possesses the CQ. On Dowe's account, there does not appear to be a difference, because possessing a CQ does not mean possessing a unique value of a CQ. On Dowe's account, processes and interactions are not mutually exclusive, and a process may undergo interactions, thereby changing the value of its CQ, possibly even possessing a value of zero without destroying the process. For example, a neutral atom possesses a charge of zero, unlike a shadow which does not possess a charge at all. However, it makes more sense to ascribe the CQ to the object, since otherwise the phrase "of an object" would be superfluous, as world lines are defined with respect to objects.

Apparently D2 means to associate one process with one CQ, but an object with more than one CQ still has the same world line. In fact, since having a zero-value CQ may still be counted as having a CQ, it appears that CQs come in packages: either an object has none, or it has them all. If the object possesses CQs, then so long as that object persists, it will continue to have those CQs and hence the same processes. In this case, we might as well say there is one process. Therefore, a process is a world line of an object, which object possesses at least one CQ.

Salmon's Definitions

I will take Salmon's definitions from his most recent article [99], which is the latest in the series of interactions between him and Dowe.

Definition 7 (S1) *A* causal interaction *is an intersection of world-lines that involves exchange of a conserved quantity.*

Definition 8 (S2) *A* causal process *is a world-line of an object that transmits a nonzero amount of a conserved quantity at each moment of its history (each spacetime point of its trajectory).*

Definition 9 (S3) *A process* transmits *a conserved quantity between A and B* $(A \neq B)$ *if and only if it possesses a fixed amount of this quantity at A and at B and at every stage of the process between A and B without any interactions in the open interval* (A, B) *that involve an exchange of that particular conserved quantity.*

The general notes made under Dowe apply here as well. For instance, world lines are really meant to be world worms. Salmon is clear about the importance of this distinction:

> The main difference between events and processes is that events are relatively localized in space and time, while processes have much greater temporal duration, and in many cases, much greater spatial extent. In space-time diagrams, events are represented by points, while processes are represented by lines. [97, p.139]

I think it will be clear that what I have been urging about the proper interpretation of "event" is precisely what Salmon means by either "process" or "interaction," once we realize that "world line" is strictly a misnomer, and that the "lines" on space-time diagrams are themselves idealizations. That settled, let us examine the details.

Definitions D1 and S1 are the same. Definitions S2 and S3 are in fact *jointly* definitions of a causal process. Salmon has defined the ideal case, where processes can be separated from interactions, and via S3 made them mutually exclusive: one ceases when the other begins. Mutual exclusivity creates a couple of interesting problems.

First, it has to be relaxed in order to be used in everyday situations. Salmon acknowledges that, for practical purposes, we can consider a thrown baseball to be a single causal process, and ignore the very large number of interactions with air molecules. Second, if mutual exclusivity were relaxed in that manner, then the very pseudo processes that S3 was supposed to exclude would invade the theory, because then there would be no principled way to distinguish transmission from mere possession, since a case of mere possession could be rendered one of transmission by relaxing the requirement of no interactions.

Third, the exclusivity requirement generates a problem of identity. Consider the simple example of a collision. Two processes lead into the collision, which is an interaction, and two leave. The world worms make a nice partial "X" on the spacetime diagram, but during the interaction, neither causal *process* exists, because the CQ in question has changed value. Every change in the CQ corresponds to a different process, and it becomes difficult if not in principle impossible to make identifications.

Fortunately, it is not required that we do so. All that is required is that the CQ is in fact conserved over the whole interacting system. The CQ theory does not require that each CQ get tagged and apportioned. Nevertheless, it is worth noting that at least for macroscopic examples, Dowe's version does not suffer the same problem. As Salmon points out, there are cases which will impose the same difficulty. Citing an example from Feynman, he wrote

> when two identical particles collide and recoil from each other, it is impossible in principle to determine which outgoing particle is genidentical with which entering particle. The concept of genidentity breaks down. [99, p.468]

The goal of this example was to show that Dowe's appeal to an idea of identity of objects (and processes) over time was not in itself an unproblematic notion. However, in neither case does the inability to made identity claims for the processes on either side of an interaction hinder the account. Observers are still free to make identity judgments on the basis of object properties. The important features for judging causality of an interaction are whether the CQs were in fact conserved. If they were not, then there was some other, unnoticed, process which entered the interaction as a source or sink.

There is more of interest lurking in S2 and S3. Dowe thought that "transmission involves no more than continuous possession," and offered to replace S2 and S3 with[5]:

Definition 10 (S2*) *A* causal process *is a world line of an object that possesses a constant amount of a non-zero CQ at each moment of its history.*

However, in making that replacement, Dowe neglected the fact that S3 also precluded any other interactions, which Salmon considered a crucial ingredient of the theory. Nevertheless, as already argued, that requirement cannot consistently be retained, and in fact our reduction of S2 and S3 to one definition will resemble Dowe's formulation. Let us see how this proceeds.

Dowe has detected a problem with the requirement of continuous possession, whether on definition S3 or S2a. The root of the problem is an ambiguity in

[5]This is Dowe's "IQ_2'" from his 1995, updated for Salmon's 1997 conversion to CQs instead of invariant Qs.

the phrase "the open interval." There are many paths between region A and region B. Since we have to pick one, the only logical choice is the path along the world-line of the process. Working from Salmon's definitions, we reformulate S3:

Definition 11 (S3a) *A process* transmits *a conserved quantity between* A *and* B $(A \neq B)$ *if and only if it possesses a fixed amount of this quantity at* A *and at* B *and at every stage of the process along the world line of that process between* A *and* B *without any interactions in the open interval* (A, B) *that involve an exchange of that particular conserved quantity.*

In addition, there is the question of whether the world line or the object is transmitting the CQ. As in Dowe's definition, the "that" is ambiguous. In Dowe's case, it did not matter which we chose. In Salmon's it does. It must be the *world-line* which does the transmitting.

If it were the object in definition S2 which does the transmission, then a causal process would simply be a "world line of an object," with some special conditions on the object. However, this is inconsistent with definition S3a (or S3). S3a says that the processes transmit the CQs, but via S2 we just defined a process to be a world line. Thus, the thing doing the transmitting must be both a temporal 3-dimensional object and an atemporal 4-dimensional object. Contradiction.

Suppose instead that definition S2 says that a causal process is a world line of an object which *world line* transmits the CQ. Obviously, by the previous paragraph, this is consistent with S3a. However, S2 and S3a are now becoming redundant. The "object" in S2 is no longer doing any useful work. And, unlike Dowe, Salmon allows that any world line can be an object, including "the patches of the surface layer of the wall that [sequentially] absorb the energy [from his rotating spotlight]," so "object" does not restrict the *kinds* of world lines that are acceptable.

Similarly, the last part of definition S2 refers to an object only incidentally. It requires that a nonzero amount of a CQ is transmitted at each moment of that object's[6] history. However, S3 already took care of that in the notion of transmission, which had to keep a fixed amount of the conserved quantity at every stage of the process. Dowe's formulation of S3a made it clear that what was meant was every stage *along the world line of that process.* Finally, although Salmon's explicit invocation of the form of the "at-at" theory is precise, it is no more so than the notion that the quantity is possessed at every stage along the world line, so we can drop the A and B notation. At this point we have,

Definition 12 (S2a) *A* causal process *is a world line that transmits a nonzero amount of a conserved quantity.*

Definition 13 (S3b) *A process* transmits *a CQ if and only if it possesses a fixed amount of this quantity at every stage of the process along the world line of that process, without any interactions that involve exchange of that particular CQ.*

Reading them together, we realize the transmitted CQ must be both a fixed and nonzero amount. Therefore:

Definition 14 (S2b) *A* causal process *is a world line which possesses a fixed nonzero amount of a CQ at every stage of the process along the world line, without any interactions that involve exchange of that particular CQ.*

[6] Here the "its" has to refer to the object. World lines do not have a history.

But a causal interaction is an intersection of world-lines which involves an exchange of a (precisely one? at least one?) CQ. But on the CQ theory, there is no way to change a CQ without a causal interaction. That is precisely the point of the theory. Although it is theoretically possible to have an interaction without the change of a CQ (if two processes exchanged equal amounts), the converse is not true. Therefore "... at every stage" is just another way of saying, "free from interactions." Therefore:

Definition 15 (S2c) *A* causal process *is a world-line which possesses a fixed nonzero amount of a CQ and is free of causal interactions which involve that particular CQ.*

We have now arrived at a formulation of S2 and S3 which can be directly compared with D2. In fact, it is D2 plus the "no interactions" requirement which terminates a causal process as soon as a relevant interaction begins. In addition, for Salmon's theory, it is helpful if each CQ has its own causal process. That way not all causal processes in a world line need to cease when an interaction involving one CQ begins. The major difference between S2c and D2 is that in D2 processes can persist through interactions. In S2c this is strictly untrue.

Comparative Evaluation

Now recall that Salmon's reason to introduce the "no interactions" requirement was to rule out pseudo processes like the spot of light circling the astrodome. To illustrate the need for the requirement, Salmon asked us to consider the "object" made of the successive illuminated patches of wall, a helix in spacetime. As the story goes, this "object," call it P, possesses a fixed nonzero quantity of a CQ—presumably some variant of energy absorbed from the light—precisely because it is made of those patches being illuminated. However, we know that it must be a pseudo process. For one, the spot can translate around the wall faster than the speed of light! Fortunately, the account excludes it because of the clause barring interactions. Although P is an object for Salmon, it is not a causal process. Hence the account has done its work by separating causal processes from pseudo processes.

Indeed, it seems to have done its work too well. The same strategy used to exclude P also seems to exclude the baseball flying through the air. A ball flying through the air interacts with the air molecules, and so as we have seen, it is not, strictly speaking, a causal process. But this is exactly the kind of example which should be a causal process. Salmon's answer is that the baseball case is just one of pragmatics:

> In many practical situations definition 3 [the transmission definition] should be considered an idealization. ... In this,[7] and many similar sorts of situations, we would simply ignore such interactions because the energy-momentum exchanges are too small to matter. Pragmatic considerations determine whether a given "process" is to be regarded as a single process or a complex network of processes and interactions. [98, p.309]

So what is the difference between this case, so easily dismissable, and the case of P? It appears that the only justification for applying the rule to the baseball is that the exchanges of the relevant CQ or CQs are "too small to matter." Yet, in baseball, the interactions matter enough to allow for curve balls. What Salmon seems to mean is that *each individual* interaction is too small to matter,

[7] Salmon's example was actually a bullet, but I am trying to support nonviolent philosophy.

and this seems to be true. The change in momentum from a single interaction is bordering on the undetectable. Is this also true of P?

If we examine the patch P without the light, it's just a gerrymandered helix of wall. Presumably the whole wall is in a consistent state, so before anything funny begins, each spot along P has the same amount of mass-energy, momentum, and charge, and there are no causal interactions. Switch on the light and the whole helix is uniformly illuminated. In its own frame of reference, P is just an unmoving patch of wall. Granted, the surface of P is always changing, but let us ignore that for now. What happens in a single interaction at P?

If the light beam is thought of as a stream of photons then P or any illuminated patch is analogous to the baseball. It is interacting with a large number of very small things, spaced very close together in space and time. Each individual photon interaction is very small indeed, and at least as ignorable as an air molecule colliding with a baseball. It does not appear that the "no interactions" rule does the job.

There *is* something strange about P, and it has to do with energy absorption. P is always in the state of first experiencing the light. Although, in its own frame of reference, P is being illuminated for quite awhile, it does not absorb the corresponding amount of energy. If the light is bright, P should be getting quite warm, but it does not. In fact, if we think about the light beam as a stream of photons, we realize that the beam intensity is smeared out along P. If the source emits n photons per second, but the spot traverses 10 times its own width in one second, then each P will receive $n/10$ photons per second. The faster the source spins, the lower the intensity of light seen by P. So in P's own frame of reference, it is sitting there being illuminated by a nonrotating beacon of intensity $n/10$. The beacon, however, is emitting with a much greater intensity. Surely this failure of conservation is an indication that at least one of our entities is not a process?

I think this insight lies behind Salmon's corollary at the end of his 1997 article. [99] After all, it was the ability to formulate this corollary persuaded Salmon to switch from invariants to CQs. Using the definition of CQ[8]:

Definition 16 (corollary) *When two or more processes [possessing a given CQ] intersect [(whether they interact or not)], the amount of that CQ in the region of intersection must equal the sum of the separate amounts possessed by the processes thus intersecting.*

For the example at hand, the light beam has a certain amount of energy, as does the patch P, but these quantities do not sum at the intersection.

3.4.2 Implications of the CQ theory

Asymmetry

The CQ theory is neutral with regard to the direction of causation. Salmon and Dowe consider this an advantage, because they want to allow that some causal influences could travel backwards in time. However, this temporal tolerance comes at a price. As it stands, CQ theory cannot account for the asymmetry of causation.

Elsewhere, Dowe has worked out a possible solution to that problem [23, 26] using a Reichenbachian-style "fork asymmetry" he calls "favoured fork theory". This is essentially a modification of the same solution Salmon had offered in an earlier version of his theory. The main advantage of the fork asymmetry

[8] I have placed apparently redundant portions in [square brackets].

account is that it makes the direction of causation an objective and contingent feature of the world. Objective because it is defined in terms of conditional probabilities. Contingent because it rests on the direction of open conjunctive forks as they actually occur in the world, and no direction is logically necessary. Nothing about the process theory itself implies a direction. A conjunctive fork is defined over three events A, B, and C such that C screens off A from B,

Figure 3.2: Conjunctive fork

which are otherwise correlated. More precisely, $\mathrm{P}(A\&B) > \mathrm{P}(A) \times \mathrm{P}(B)$, and $\mathrm{P}(A\&B|C) = \mathrm{P}(A|C)\times \mathrm{P}(B|C)$. As indicated in the drawing, this probabilistic condition specifies nothing about the direction of the conjunctive fork. It is possible to find forks where C occurs after A and B. All that is needed is for a common condition to result in a common fate. Then, this fate F screens off A from B. As long as we can find an A and B which are correlated, the triad AFB constitutes a "backwards" conjunctive fork. But this is easy: take the A and B which resulted from C! In that case, we have a *closed* fork: The fork asymmetry thesis is just that all or most *open* forks are open to the

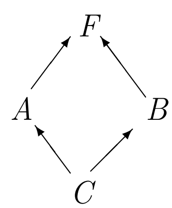


Figure 3.3: Closed fork

future. The way to use this in giving asymmetry to causal processes is to equate the direction of a causal process with the direction of any open forks in which it is involved, or the average direction of open forks within the larger network in which it is embedded. Dowe's solution is to combine both options, with the former criterion used first where available, and the second only as a backup in cases which would otherwise be left without a causal direction. Dowe wants the individual definition of direction to be primary in order to allow some processes—he has in mind a particular solution to or explanation of the Bell effect—to travel backwards in time.

"Time" could be given by some other notion, or more conveniently, just defined by the overall direction of the whole network, which is to say the direction of the (majority of) the open forks in that network. On the latter reading, the fork asymmetry thesis is that that most open forks are open in the same direction. That direction is called the future. So, if a process L is involved in an open fork, then it is defined as pointing in the direction of that open fork, regardless the direction of the other processes around it. If the direction does not match the majority, then L is a backwards-in-time process. Dowe does not consider the status of L if it is involved in open forks in both directions, if such exist. Presumably its direction would still be undefined, but so long as causal direction is defined for most processes, the theory has done its job. Although the theory is at least partially extrinsic, that in itself is not an objection, but a possibility in line with Mach's speculation that inertia arises from the total distribution of mass-energy in the universe.

For our purposes, a theory which allows for backwards causation is not a problem so long as it does not insist upon it. Since the fork theory leaves this matter contingent upon physical reality, it will serve our needs fine so long as it provides for the usual asymmetry, which it does. In fact, it exhibits a very nice, if intuitive, compatibility with entropy. When we execute a break in billiards, we start an irreversible process. Even if we had an ideal billiards set which lost no energy, the event where the cue ball breaks the set is different from the reverse event where the scattered balls form the set and eject the cue ball. The difference is that in the first, the collision with the cue ball screens off correlations among the other balls in the future, while in the second event, the collision with the cue ball screens off the past correlations among the other balls. The open forks point towards the direction of increase of entropy in the local system. If time is defined externally or by reference to the broader network, then any local decrease in entropy will involve forks open to the past, and so constitute a backwards-in-time event. Likewise, if we now add in thermodynamic entropy, the dissipation of macroscopic energy to heat, such a reversal becomes even more unlikely, though theoretically possible. But now even simpler cases constitute backwards causation. A single cue ball starting from rest and beginning to accelerate involves the unimaginable correlation of microscopic movements leading to a common future effect (the rolling) which screens off those movements: a fork open to the past.

On the other hand, because the microscopic causes conspiring to produce this movement are by definition invisible, the phenomenological description of the motion is that of an apparently uncaused acceleration, which is equivalent to an unseen injection of energy. Either such an unseen increase in energy is due to a real unseen cause, like a wind or a magnet, in which case there is no fork open to the past, or it is not. In either case the appearance is the same: an increase of energy with no visible correlate. That ought to look strange. Even if there *is* a hidden cause, it will *appear* to be a case of backwards causation, unless some cause can be found or inferred from the particular nature of the motion. The fact that even movies run in reverse look so peculiar indicates that it does not take much to overwhelm any inferential ability to extract a cause. However, another fact of perception thickens the plot. Minimalist movies run in reverse do not look like cases of reversed time. They look like animate motion. [9]

If the perceptual system treated animals as collections of billiards balls, then animate motion would appear to be backwards causation. It does not. In fact, this system "kicks in" to resolve certain apparent cases of backwards causation, by attributing animacy. But more importantly, animals get parsed as single causal processes (objects), and so there is never a question of performing raw correlations between the movements of different limbs. And as soon as the organism is embedded in a surrounding, it becomes analogous to the cue ball, that is, the common cause of a series of correlated effects in its vicinity. That is why movies have to be minimalist in order to evoke notions of animacy, and why running a breaking shot in billiards in reverse would just look like what it was: a case of backwards causation.

In the psychology chapters it will be important to remember the close connection between causation, direction, and entropy, particularly at the perceptual level, since perception is sensitive to causal direction. My project is to formulate a notion of causation which both underwrites and makes sense of causal perception. Because I think the CQ account is the best available account which fits the psychological evidence, I have developed some experiments which test the effect of the manipulation of CQs on the perception of causality, indirectly by testing how those manipulations affect naturalness. Despite my focus on energy and momentum manipulations, what I found was that the apparent thermodynamic entropy resulting from my manipulations was at least as fundamental to the perception of naturalness as the conservation of classical CQs. What I have hoped to show in this section is that the CQ theory can handle those results quite well, and in fact provides a natural explanation for why unexplained anti-entropic motion looks strange: it is equivalent to backwards-in-time causation.

Hume's Ghost: Reprise

Having worked our way through the details of the definitions of CQ theory, where have we arrived? The primary goal of the account was to separate causal processes from pseudo processes, allowing causality to be analyzed in terms of processes and interactions, hopefully thereby avoiding Hume's difficulty, and providing material for causal explanations. Salmon and Dowe formulated the account to meet the following requirements (summarized in [24]). The account needed to be:

1. objective
2. contingent
3. compatible with indeterminism
4. time-independent
5. free of Humean hidden powers

CQ theory is objective in that it refers only to world lines and conserved quantities. It is contingent in that it gives an account of physical causality in the world, and is contingent upon features of the world such as the existence of CQs. It is not a logical account of causation in all possible worlds. Although we have not yet discussed the point, the theory is compatible with indeterminism. All it requires is conservation, and no change in processes without interactions. A spontaneous decay of a radioactive atom is an intersection (of the Y type), and obeys conservation laws. Since the theory refers only to spacetime regions and does not link causation to the direction of time, it is time-independent. As we just saw, it is quite independent of the direction of time. However, is it free of "spooky action at adjacency"?

Salmon formulated his process theory as a direct answer to Hume's challenge. Hume could not see any causal connections between events. Salmon replied that it was an error to believe it was possible to separate the world into "causes" or "necessary connections" separate from events. In addition, he traced part of Hume's problem to his Zeno-inspired discretization of events. In particular, by defining transmission in terms of an at-at theory[9], he had avoided the infinite regress of causal relations among the constituents of the process. More simply, a process is already an extended, continuous entity, so Salmon never has to attempt to build up the continuum out of discrete points.[10] As we saw in S3,

[9]A process transmits a mark (or CQ) from A to B iff it has the mark (CQ) *at* A and *at* B, and *at* every point in between, period. There is never a question of "getting" from A to B, only of being at a certain point at a certain time.

[10]Isaac Barrow very laid out this answer to Zeno, Epicurus, and other atomists very clearly in lectures given at Cambridge in 1665 and published several times thereafter. [3] We know Hume read Barrow. It is a shame he did not pay more attention.

Salmon carried the at-at theory into his formulation of the CQ theory, certainly on account of his conviction that "With the aid of the 'at-at' theory, we have a complete answer to Hume's penetrating question about the nature of causal connections."[97, p.157] By restricting the theory to extended portions of world lines in space and time, the same thing is accomplished. First, in a spacetime diagram, there is no "getting" anyway. Second, by disallowing instantaneous time slices to count as events, we avoid the problem of connecting one event to the "next".

However, according to Dowe, Salmon's original process theory and the CQ reformulation involve "hidden powers" of precisely the sort Hume would abjure. There are two candidates for hidden powers: counterfactuals and propensities. Salmon's original theory had to resort to counterfactuals in order to resolve problems about pseudo processes. The CQ theory does not need them. However, in order to handle cases of indeterministic causation it posits that causal processes have *propensities* to produce different effects. These are construed as real things: tendencies, dispositions, or chances. As Dowe summarized, "propensities are probabilistic causes, the fully objective, relational properties of causal processes. They are not directly observable."[24] A classic example of a truly indeterministic propensity is the decay of a radioactive atom. Over a duration equal to the atom's half-life, it has a chance of decaying equal to 0.5.

In accordance with Humphreys [58], Dowe wants to interpret this as an objective, indeterministic causal process with a propensity of 0.5, and Salmon appears to agree [98, p.301]. In fact, this would seem to be the only tenable interpretation since the process theory is such a singularist account:

> causal processes, in many instances, constitute the causal connections between cause and effect. A causal process is an individual entity, and such entities transmit causal influence. An individual process can sustain a causal connection between an individual cause and an individual effect. Statements about such relations need not be construed as disguised generalizations. [97, p.182]

Although propensities are not observable in a way that would satisfy Hume, they do have observable consequences. In this respect they are not entirely different from CQs like mass-energy, momentum, and especially charge. What is the status of CQs?

CQs either involve the idea of conservation laws, or the bald assertion that as a matter of fact, some quantity *is* conserved, and so can serve as the appropriate "mark". In either case, for the theory to be adequate it must assume the truth of something which cannot be established on purely empirical grounds. That would not have satisfied Hume's general skepticism.[11] On the other hand, it would be a mistake to try to offer an account of physical reality which did satisfy Hume's general skeptical argument. Hume's point is entirely sound: empirical knowledge cannot be logically necessary. Dowe and Salmon both accepted this thesis when they chose to formulate a *contingent* account of causality, and that was a good move. There may be no single causal account valid in all possible worlds, which would imply that several approaches are doomed from the start. Even if there is such a single, logical definition, it is likely to be so broad as to be unhelpful; science is more interested in the causality that affects the world we inhabit.

Salmon thinks the CQ account requires only bald assertions of conservation, and does not involve any laws.

[11] On the distinction between Hume's general skepticism and his less skeptical rules of reasoning in philosophy, see De Pierris [22].

> We might be tempted to say that conserved quantities are those quantities governed by conservation laws, where by "law" we mean either a true lawlike statement or a lawful regularity in nature. If we were to take this tack, however, we would free ourselves from the curse of counterfactuals only at the price of taking on the problem of laws. . . . This may not take us out of the frying pan into the fire, but it does seem to offer another hot skillet in exchange for the frying pan. The hazard can be avoided, I think, by saying that a conserved quantity is a quantity that does not change. [98, p.310]

Given that the CQ is based on interactions whose definition involves changing CQs, what did Salmon have in mind in that last sentence? He continued, "I am prepared to assert that the charge of the electron is 4.803×10^{-10} esu, and that the value is constant."

The intent cannot be to limit CQs to fundamental constants of the universe. Granted, those do not change (very fast), but that is not the kind of property discussed in conjunction with "manifests a conserved quantity." There, Dowe and Salmon talked about momentum, energy, total charge, and the like. These things change all the time. Salmon must mean one of two things here. The first possibility is that a CQ is a quantity whose total value, for the whole universe, does not change over time. The second is that a CQ is a quantity whose total value for the whole system of interacting processes is the same on both "sides" of the interaction(s).

The second definition is more empirical, but it is insufficient. An interaction may happen to leave a quantity unchanged even though that quantity is not normally considered a CQ. Color, temperature, volume, weight, and other properties suggest themselves. Trying to patch this definition by adding "always" or "for all interactions" is just to turn it into a conservation law, which is precisely what Salmon wanted to avoid. What is needed to define a CQ at this level is a sense that the CQ *cannot* change, rather than the possibly accidental fact that it *does* not.[12]

What about the possibility of definining a CQ based on what in fact does not change on the grand scale? According to that definition, if charge is conserved in all interactions, then the total charge in the universe now is the same as it was when the universe began, and will remain the same, at least modulo vacuum-point fluctuations. This formulation amounts to a plain universal generalization: "As a matter of fact, it happens that the total charge in the whole universe is always constant." That statement, plus the assertion that CQs can only change through local interactions, would yield the desired charge conservation.

However, it will probably also yield a host of accidental CQs as well, especially if we allow Goodman-esque gerrymandered properties. In our universe, parity is not quite conserved. What if as a matter of fact, parity *happened* never to change? What about socioeconomic concepts like "average purchasing power per capita"? The underlying question is this: why should causality be tied to quantities that just happen to be conserved? The appeal of the theory is that CQs are fundamental features of physical reality. What we want are quantities whose conservation is *impossible* to violate, at least without substantially altering the structure and workings of the universe. Salmon's move avoids laws at the price of undermining the motivation for using CQs in the first place. If CQs are not somehow truly fundamental, but just happenstance conservations, there is no motivation for grounding causality in conservation.

[12] I thank Michael Friedman for insights on the dependency of CQ theory on laws.

It is also important to consider the notion of correspondence. The true CQs may be very abstract, complex entities. My astronomer friends caution me that mass-energy is not globally conserved in general relativity, although it is locally conserved. Because the structure of spacetime is differentially curved, the integral between points depends upon the path taken. Clearly, if the CQ theory is to link to our human notions of causation and our ability to perceive causal interactions, we cannot mean that causal notions are tied to partial differentials of some general relativistic tensor. Indeed, even "mass-energy" is stretching things. We do not refer to mass-energy when discussing baseball. There is no real problem here. Just as Newtonian—or dare I say, occasionally Aristotelian—physics really is the physics of everyday events from golf through moon landings, so too there is a correspondence between ultimate CQs and those conserved at the level of everyday experience. Differential conservation of mass-energy nevertheless is global within flat spacetime, when the integral is path-independent.[13]

What about the idea of propensity? In dropping the idea of necessary connection, the theory has adopted "propensity," which Dowe considers to be problematic, if essential. With genuine indeterministic processes like radioactive decay, that does not seem to be a problem. As a consequence of the structure of the atom and the laws of the universe, it has a certain chance of decaying. That may not be very satisfying, but it appears to be true. And even in the everyday world, as Hume observed, our experience is probabilistic, even were one able to maintain underlying determinism. Further, so far as the consequences of having a certain propensity are testable[14], they are empirical claims. In fact they are dispositional properties. Both a propensity to decay and a disposition to dissolve arise from the structure of the substance in question, and manifest themselves in behavior.

I think that Dowe's charge of hidden powers is misplaced. The CQ theory should not be required to meet Hume's conditions on causality, for I have argued that they cannot be met. Hume defined the problem in such a way as to make it insoluble. The process theory did not "solve" Hume's problem in a straightforward way: Salmon's theory did not point to some extra something which was the "necessary connection." Instead he adopted the ontology of (potentially) continuous processes in spacetime and argued that these dissolved the problem of "connectability." The CQ theory took the independent step of adopting certain laws as fundamental, and said that within that context, it was possible to say what kinds of connections counted as causal.

One of the reasons process theories in general are well-suited to answering Hume is that they are not bound by the same space-time ontology which Hume possessed. Hume had an essentially discrete view of the world backed by an atomistic understanding of space, time, and mathematics.[15] This led him to parse the world into cinematic moments and then wonder how they were connected. Salmon's answer is, essentially, "you can't do that." Hume was not the last philosopher to make that kind of mistake, and part of the point of Russell's chiding in 1912 was to bring the ideas of dense or continuous sets into the debate. In 1984, Salmon still felt it necessary to argue against point-event theories.

[13] Thanks to Robert Link (personal communication) for pointing out some basic features of general relativity.

[14] I would argue that they are no less testable than other empirical claims. True, a propensity of 1 is compatible with any observed sequence, but the degree of belief one should have in the hypothesis decreases with every disconfirmatory observation.

[15] For excellent discussions which do credit to Hume's ideas and their sources, see Marina Frasca-Spada [35, 36]

Paul Humphreys has put the point very clearly:

> within the logical and mathematical spaces inhabited by the abstract representations of cause and effect, operations are possible that simply cannot be done via real manipulations . . . in the world because of its causal structure. Yet by dealing with observed relative frequencies and mathematical operations upon them, one is dealing with (abstract representations of) observed effects of the world on us, and these observations are independent . . . in a way that their origins, the actual causes and effects, are not—one can separate the observations in conceptual or mathematical space in a way that is not always possible in the world itself. Indeed, this is exactly what Hume did in insisting that cause and effect are 'distinct existences', and this construal in terms of logical independence, rather than causal independence . . . moves one in a direction that makes a successful characterization of causal relations, as opposed to noncausal relations, very difficult. [58, pp.51-52]

In short, Hume was convinced—in part because of his ontology of space and time—that he could analyze and separate things which in fact he could not, except logically. As we have noted repeatedly, that was both appropriate and successful for refuting the rationalist idea of causation as necessary connection. But it was a mistake to carry over the same methods into a discussion of empirical causality.

Michael Friedman has pointed out (personal conversation) that the CQ theory could be formulated on a discrete spacetime. The trick to answering Hume is still the same, only it is harder to motivate. So long as one does not introduce another level of causality between adjacent moments in spacetime, then there is never a question of how to "get" from point 1 to point 2. A process going from 1 to 4 is simply *at* point 1 at t, at 2 at $t + 1$, at 3 at $t + 2$, and at 4 at $t + 3$. Possession and exchange of conserved quantities follow naturally. The conceptual difficulty with this ontology is that if events resolve to finite, discrete units, it is *natural* to see them as separate, essentially disjoint movie frames, and to ask Hume's question. I think the solution is anticipated in the practice of event perception, which often has to show observers events which are in fact simulated on digital computers with discrete frames. Events are necessarily shown as multiple-frame events. Velocity cannot be determined or understood except across two frames, and acceleration requires a minimum of three. However, it is a mistake to assume therefore that velocity is derived from a computational comparison of frame-based positions.

Hume was led to deny continuity and denseness by his strong empiricism, as shown in his discussions on infinites and infinitesimals. Nevertheless, his overall empirical project was to advance the sorts of claims made by Salmon and Humphreys: that causation is not a logical phenomenon, but an empirical one, both contingent and uncertain. Unfortunately, the combination of his discrete ontology, his inadequate idea of what kinds of connections were possible, and his insistence that causation still required necessary connections, of the logical sort, he concluded that causal connections did not exist in the world, but only in the mind.

So yes, the CQ account does offer an alternative to Hume's constant-conjunction theory, precisely because it can point to causal processes, as primitives, and say "here are the connections you seek." But it does not do this in Hume's scheme, but in a modified ontology. It does not answer Hume's general skepticism, nor should it try, but it does provide an account "within natural philosophy," even if not playing entirely by the rules of Hume's unwinnable game.

Evaluation

We have here exactly the remedy that we were looking for to supplement Ducasse's account. In addition, it is the kind of theory of causality which can underwrite causal theories of knowledge [44], which were formulated to overcome certain paradoxes in pure logical epistemology.[16] The CQ theory specifies which kinds of processes are causal. They may not completely match the processes *perceived* as causal, and more on that later, but CQ theory can serve as a guide.

So, granting that CQs do exist, the CQ theory ties causality to objective features of the world. "Show me a causal interaction," says the CQ theory, "and I will show you the exchange of at least one CQ." Causal interactions and causal processes are defined in terms of objective physical quantities, and so are physical entities. The answer to Hume is that causal connections are causal processes, which is to say physical instead of logical connections. The CQ theory, then, offers a potential answer to Hume's *ontological* worries. However, it does not directly address the issue of causal *perception*. Hume wanted also to know about how observers could come to know about causes. On his explanation this was easy: causes existed only in the mind anyway, so it was easy to know them. Salmon and Dowe have given a more satisfactory notion of what a causal connection *is*, but how do we know it?

Part of the answer is exceedingly simple. If a baseball in flight is an example of a causal process, we just look at it! We know causal processes because they are around us all the time. This response has some truth to it, but it misses the deeper part of the question. Perceptually, empirically, how do we separate out the causal from the noncausal processes? The CQ theory makes it easy: causal processes are those which manifest a genuine CQ. But how do observers know what the CQs are? Or do we?

The fact that the CQ theory is so appealing means both that it connects to physical theories and that it intersects strongly with our intuitions about what is and is not a cause. Intuitions are important in philosophy, perhaps overly so, and it is hard to avoid them. How are theories of causality tested? Against our scientific understanding of the world, yes, but also against intuitive notions of causation, which is to say doubtless cases of cause and effect, or the absence thereof. The purpose of philosophical analysis of causation is to show how an apparently essential concept can be made to tie in to our best theories of the world. There is a fairly standard gauntlet through which a theory must run: shadows must not be said to cast flagpoles even if the equations are reversible, hexing the salt must not be said to be the cause of its dissolving, untreated syphilis must be the cause of paresis, even if both the conditional and the unconditional probabilities are low, and barometers must not be said to cause weather changes just because they always covary with and precede them. All of these examples came from appeal to intuition, which means that our intuitions, however acquired, are in part responsible for our theories of causation.

If CQ theory is right, then our intuitions must somehow have preferably latched on to causal processes or causal interactions, rather than their noncausal counterparts. The questions are: "how?" and "how accurately?" What exactly

[16] I find it interesting that Goldman sought a causal theory of knowledge to escape from logical theories: "The analysis presented here flies in the face of a well-established tradition in epistemology, the view that epistemological questions are questions of logic or justification, not causal or genetic questions." It is of course precisely the failure of purely logical analyses of causality which led the the formulation of the process view.

is available to observers and when? How does science connect with common sense? Since CQ theory appeals to objective features of the world, by definition those features are *potentially* available to perception, or at least to inferential judgment. Observers need not be sensitive to actual CQs, but the success of CQ theory would predict that at minimum observers can pick up on some correlate of at least some CQs. Natural selection and evolution would predict that organisms would be more attuned to information which is more relevant for survival.[17] It is not obvious what those would be, although presumably energy and momentum make good candidates.

In other words, the CQ theory supports the possibility of causal perception (or judgment), because it links causation to objective features of the world. There is no guarantee that observers can pick up on those particular relations, but the success with which CQ theory meets our intuitive ideas of causation suggests there is at least some approximate match. If some causal interactions are perceivable, then we have an answer to Hume's *psychological* challenge as well as his ontological challenge. We know it is not the case that all causes are directly knowable: some are completely beyond our native sensory modalities. Yet intuitively, some causes are known directly and immediately, at least against the background of a developed perceptual system. In fact, since the CQ theory identifies some very ordinary and perceivable events (baseballs, bats, and the like) as causal processes, there is every reason to believe that *some* causal processes and interactions can be perceived directly or inferred from the data of perception.

But which events are in fact identified as causal? According to what properties? Are there some noncausal events which are judged causal? Are there some causal events which are judged noncausal? Are there physical or perceptual properties which can predict observers' attributions of causality, and do these properties correspond in any way to CQs? What *are* peoples' experiences of causality, in general?

These questions about causality and causal attribution are central to the debate because they inform us about human knowledge of causality, which is the well from which the debate springs. They are empirical questions, best pursued in a psychological context. Philosophers can gradually work out intuitions, but the resulting intellectual ping-pong has no ultimate resolution so long as it is my opinion against yours. Not that democracy ever resolved anything either, but an appeal to actual human performance would keep the debate honest, and give philosophers more to talk about than theories and personal intuitions. For instance, perhaps momentum and energy are equally relevant objectively. What does it tell us if observers can reliably detect and understand one CQ, say energy, in certain visual circumstances, but in similar circumstances fail to use momentum? Or what if observers are sensitive to nonconserved quantities, like kinetic energy, or even entropy? What if different observers lock on to different CQs? Would answers to any of these suggest a need to revise our theory of causality? Indeed, given the process and interaction theory, is it possible to formulate the theory in terms of exchanges, losses, or gains of *non*conserved quantities? What if observers are uniquely sensitive to gain or (non-loss) of entropy and are insensitive to CQs?

[17]Using the converse of the anthropic principle, we can infer that charge detection has not, statistically, been important for survival. On the other hand, the fact that we can do it at all suggests that it may not have been wholly irrelevant. More likely this is one of Gould's spandrels: natural or sexual selection for quantity of body hair with the unselected corollary that it is possible to detect a charge buildup on the skin. Given that spandrel however, selection could have begun to work on molding humans who react appropriately when about to be struck by lightning.

Chapter 4

Psychological Approaches

> For to me it seems evident, that the essence of the mind being equally unknown to us with that of external bodies, it must be equally impossible to form any notion of its powers and qualities otherwise than from careful and exact experiments, and the observation of those particular effects, which result from its different circumstances and situations.
> *Treatise*, p.xvii

Events are extended in time and possess intrinsic dynamics grounded in conservation laws. That observers can recognize and react to event types is possible because the underlying dynamics of different types of events dictate very specific kinds of motions—kinematics. Observers' ability to recognize and react appropriately to events is well-documented, as is their inability to articulate or reason about the "knowledge" implicit in their behavior. In this way, Hume was right: basic knowledge of cause and effect is more a matter of habit than reason. However, as we saw in the last chapter, not every regularity is causal. Furthermore, neither do observers perceive every regularity as causal. Therefore, it is worthwhile to investigate the conditions required for the perception of causality. The psychological foundation of causation is event perception.

Event perception is to be contrasted with investigations of explicit reasoning and judgment. Such cases also involve causal reasoning, but they all presume a background knowledge about the kinds of causes operating in the world. It is my contention that this background knowledge comes from event perception, and that this is the reason that Hume's billiards table provides the most basic example people can find to illustrate a clear case of cause and effect. It is these cases which provide the underlying idea of 'mechanism' invoked in the studies by Ahn [2, 1]. In turn, the kinds of events picked out as causal interactions, and the reason for their primacy, can be understood in light of the CQ theory of causation developed in chapter 3.

The present chapter presents the psychological investigation of causal perception, starting of course with Hume. It then turns to the animations of mid-twentieth-century French psychologist Albert Michotte's, which are the starting point for current research in event perception. The field of event perception is then developed with attention to the perception of causation, and contrasted with work in the field of naïve physics. Absent from this chapter is the substantial body of cognitive research into deliberative causal judgment and causal reasoning pursued by psychologists such as Ahn, Cheng, Holyoak, and Spellman [2, 1, 17, 16, 105], which is perhaps underwritten by perceptual causation, but is not directly relevant to the line of experiments developed in chapter 5.

4.1 Hume as psychologist

Hume's skeptical arguments against knowledge of causes were epistemological, and might plausibly be confined to philosophical justification. However, not all of Hume's program was pure philosophy. According to the subtitle of the *Treatise*, his project was "An Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects." Hume took himself to be applying the tools of the natural sciences to the moral, or human sciences. Stroud, for one, took such passages seriously and argued that we must understand Hume as "putting forward a general theory of human nature in just the way that, say, Freud or Marx did." [107, p.4] In short, after showing the inadequacy of the rationalist program, Hume set about to understand human nature from a point of view that we might call psychological.

It is unsurprising then that modern psychologists as well as philosophers take Hume as a starting point. Hume the psychologist posed at a more conceptual level the classic problem of the under-determination of the world by perceptions. Where Descartes worried about the perception of the three-dimensional world based on two-dimensional retinal images, Hume puzzled over how our minds "inferred" causes from mere regularity.

> It appears that, in single instances of the operation of bodies we never can, by our utmost scrutiny, discover anything but one event following another.... So that, upon the whole, there appears not, throughout all nature, any one instance of connexion, which is conceivable by us. All events seem entirely loose and separate. One event follows another; but we never can observe any tye between them. They seem *conjoined,* but never *connected*. [55, Sec.VII, pt.ii]

According to Hume, single events do not create expectations. Michotte took this passage as a denial of the existence of a singular causal impression altogether, and concluded that Hume had overstated his case. In fact, Hume did not deny that we have causal impressions, but he thought it was not inherent in objects or events themselves, but only in their repetition. Taken by itself, the statement, "All events seem entirely loose and separate," is flatly wrong. Events in the world do *seem* connected, even if in reality there is only constant conjunction. But Hume was only denying that we could have had such impressions from single unrelated observations. Nevertheless, this is not Hume at his most careful, and it is not surprising that he has been interpreted more extremely.

From such passages, Michotte [80] concluded that Hume was committed to deny not only the justification for causal inference, but the phenomenal impression of causality itself. In Michotte's defense, he was justified insofar as Hume did deny *singular* causal impressions, at least in most cases. In an extensive set of experiments Michotte evoked causal impressions from single experiences, and argued that whatever the status of Hume's philosophical thesis[1], his psychological claims were mistaken. Michotte's reading has been influential both because it fit the empiricist sketch of Hume, and because it provided a convenient cartoon of Hume in contrast with which, psychology showed clear advances.

It is, however, important to realize that in many places Hume explicitly acknowledged the impression of causality, despite appearances to the contrary in passages like the one Michotte found so objectionable. For example:

> *When I see a billiard ball moving toward another, my mind is immediately carried by habit to the usual effect, and anticipates my sight*

[1] By which he meant Hume's destructive argument against rationalism.

> *by conceiving the second ball in motion. . . .* I also *believe* that it will move. [54, p.189]

> . . . nor is it by any process of reasoning he is engaged to draw this inference; but still he finds himself determined to draw it, and though he should be convinced that his understanding has no part in the operation, he would nevertheless continue in the same course of thinking. [55, , Sec.V, pt.I]

In the midst of defining causes in the *Inquiry*, he also stated, "The appearance of a cause always conveys the mind, by a customary transition, to the idea of the effect." (p.87)

In addition, Hume had another way to defend his claim. In effect, he could assert that Michotte never succeeded in producing a singular impression, because any display which seemed to evoke a feeling of necessary connection did so only because it was sufficiently similar to the lifetime of experience built up from observing the world.

The debate has become quite involved in the literature of event perception. Gunnar Johansson, a gestalt psychologist famous for his studies in dynamical perception, set out the psychological challenge:

> I previously raised the somewhat provocative question of whether stimuli for perception of dynamic aspects of events existed and thus whether knowledge about forces acting in an event is really immediate perceptual knowledge. Most of my readers know that this was, in fact, Hume's famous problem regarding the perception of causality. We know that Hume's answer was in the negative. However, we also know that Michotte, in his pioneering investigation on causality perception (1946/1963), experimentally attacked this problem—and refuted Hume's deduction. [118, p.43]

Johansson cast Hume's point as an empirical matter, and not a deduction. As an empirical argument, it was falsifiable. Yet his target was not, as he wrote, Hume's skeptical arguments against rationalism, which are deductive. Instead, Johansson and Michotte's critiques targeted the constructive claims made by Hume in the passage just cited from Sec.VII. Hume had argued that there was no singular causal perception, and hence no singular causation. Johansson and Michotte claimed that forces and other dynamic quantities were available to perception, even in single events, although they would have to allow that the observer who could extract such information must have a great deal of experience in the world. The scaffolding Hume left behind appeared to be inadequate to construct a replacement for what he dismantled, so Johansson and Michotte abandoned Hume's foundation and started by affirming the reality of singular causation.

4.2 Michotte

> The results of the research described in this book seem difficult to reconcile with the many different theories put forward by psychologists and philosophers as to the origin of the idea of causality and its application to the data of experience. Michotte 1963

Around mid-century French psychologist Albert Michotte began to investigate the perception of causality, and concluded that Hume had gone too far. Michotte argued that *causal impressions* were real, immediate, and innate responses to single events. Michotte devised several simple experiments to investigate human perceptions of causality. His explanatory theory has not survived intact, and for good reasons, but experimenters continue to expand upon his empirical work.

4.2.1 Michotte's experiments

In Michotte's standard experiment (see figure 4.1), the observer sees two squares, A and B. Square A moves towards B, stops when it just touches B, and then B moves in the direction A was moving. If the speed ratio v_A/v_B, and the delay between adjacency and subsequent motion fell within certain broad constraints, observers perceived the display as a collision between A and B, where the collision with A was the cause of the motion of B. Otherwise they saw this "launching" impression was absent. Michotte dubbed the "illusion" of causality in this display the "launching effect." Similarly, sometimes after the collision A and B went off together, giving the impression that A carried B along, something he called the "Entraining Effect." There were other variations as well, such as launching at a distance, orthogonal launching, and changing other properties like color or visibility.

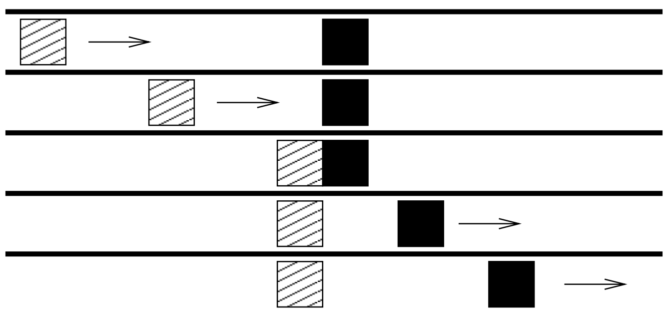


Figure 4.1: Schematic illustration of Michotte's standard experiment: the launching effect. Observers see the hashed square A begin moving to the right, and impact the dark square B, whereupon the dark square begins to move to the right. Redrawn from [109, p.69]

The launching effect works best when the incoming velocity of A is about 5-10 times the departing velocity of B. There are also timing concerns. A delay of up to about 50ms was tolerated between the collision and the departure of B. For delays of more than 50ms, direct launching was no longer constantly reported. Up through about 100ms, participants still occasionally reported seeing the event as a launching event, but by 150-200ms, participants saw the movements as independent events or triggering events. Similar constraints held for the entraining effect.

If the speed ratio was such that $v_B > v_A$, then participants saw a *triggering* event: the arrival of A triggered the departure of B. Square A may have been a cause, but it was not a sufficient cause for the resulting velocity, despite (a) the perfect correlation (Hume), and (b) the fact that there was no other change (Ducasse). Michotte explored many other constraints as well. Among the most notable, he found that some gap is well-tolerated so long as there is no delay, and that in-line collisions elicited causal responses more readily than skew collision.

Michotte also demonstrated paradoxical cases of perceived causality, although not with the same reliability as more realistic causal scenarios. In these settings, sequences which blatantly defy experience are perceived as causal. For example,

> it is even possible to produce the paradoxical case where the movement of B is slowed down by the impact of an object of the same apparent mass traveling more rapidly [in the same direction]. Thus when A moves at 30 cm. per sec., and B moves at 15 cm. per sec. before the impact but continues its movement at a speed of 7.5 cm. per sec. after A comes to a halt, a perfect launching impression can be obtained. [80, p.71]

Michotte does not report the methods or results of this experiment in any detail, but it is surprising. Michotte thought it counted against Hume's theory, since participants had causal impressions clearly not derived from past experience. On the other hand, since Michotte held that only "push" effects would seem causal, this is paradoxical for him as well. Indeed, it goes against his previous results which had established constraints on allowable velocity ratios. Furthermore, what Michotte meant by "same apparent mass" was "same size on screen".

This case is in effect similar to the "triggering" cases, where A clearly appears to be a causal element, in the Ducassian fashion of being the only perceived change before the change that is the effect, and which elicits an impression of some hidden cause. Runeson [94], made essentially the same suggestion in this middle of his extended analysis of the dynamical representativeness of Michotte's experiments. Overall, Runeson found that Michotte sampled a very small and specialized region of collision space, and was hampered by a limited notion of causality which did not allow the impression of attractive causes, or the inference of different masses based on the kinematics. There was another branch to Michotte's experiments which are worth mentioning, however, the qualitative causality experiments.

In the qualitative experiments, the "movement of one object is linked with qualitative change in another," such as color. Although it was possible for a causal impression to occur, in general it did not, even though within the scope of the experiment the three Humean and even Ducassian conditions were met. [80, p.231]

In Michotte's experiment 73, only B was shown, stationary in the middle. Suddenly, A appeared next to B, and B began moving. In this case, about half of the participants reported a launching event, the rest merely saw temporal contiguity. Over the next several experiments Michotte let A move and strike B, at which point B changed to A's color, or disappeared entirely. Again, about half of the participants reported this as a launching effect. Reversing the experiment, Michotte then began the display with A and B adjacent on-screen. A then either disappeared or changed color, at the same time B began to move. In this case there were no reports of launching, and only one or two weak impressions of triggering. Almost all of the participants reported only impressions of temporal contiguity.

Michotte's experiments should be evaluated cautiously. His methodology consisted exclusively of qualitative free-response, which is fraught with problems, and he selected his participant pool to exclude those who could never get causal impressions even in the 'obvious' cases. Overall, however, the experiments indicate the importance of *sensory modalities* in causal perception. Humans are much better at connecting some kinds of properties than others, regardless of constant conjunction. In particular, the impressions of causality in cross-modality experiments were far less reliable or strong than in the cases limited to motion. In cross-modality cases, sudden appearance could be seen as a cause of motion, but sudden disappearance could not. Taken as a whole, the the experiments display a pattern which confirms suspicions that neither Hume nor Ducasse had the correct, or at least complete, story.

4.2.2 Michotte on Hume

Michotte believed that Hume denied the existence of the "causal impression."

> It seems certain that Hume did not realise that there was such a thing as the *causal impression.* His writings are so definite on this point that the matter does not admit of any doubt. [80, p.255]

At first pass this is somewhat puzzling. After all, was not Hume engaged in explaining that very thing he is said to have denied? Michotte seems clearly to have been in error. But "causal impression" was for Michotte a technical word, which he kindly described for us in a footnote. It signified "the idea of an immediate datum, of something directly 'lived"' [80, p.15,n.20], and so his theory was that,

> certain physical events give an immediate causal impression, and that one can 'see' an object *act* on another object, *produce* in it certain changes, and *modify* it in one way or another.

The emphasis appears to be on "immediate." Michotte was arguing that the causal impression is direct, primitive, singular, immediate, and 'seen' rather than 'inferred' through repeated experience. Even this thesis could collide only obliquely with Hume's account, depending on whether Michotte focussed on the origin of the causal impression or its operation in daily life. It becomes clear in reading Michotte that he believed completely novel causes could be perceived in singular cases, as in Ducasse's example of the mysterious package.

Michotte offered several examples of what kinds of events and impressions he meant:

> that of a hammer driving a nail into a plank, and that of a knife cutting a slice of bread. The question that arises is this: when we observe these operations, is our perception limited to the impression of two movements spatially and temporally co-ordinated, such as the advance of the knife and the cutting of the bread? Or rather do we directly perceive the action as such — do we see the knife actually cut the bread? The answer does not seem to me to admit of any doubt. [80, p.15]

Hume thought that careful reflection revealed only the correlation, but Michotte denied that we could make the separation Hume claimed we could. Michotte was sure that observers saw the driving and cutting. There is a tremendous amount of specificity in the conjoined events. Michotte claimed that the perceptual information was enough to *specify* causality, presumably even on a first impression.[2] On this reading, Michotte understood Hume correctly, and flatly rejected Hume's arguments against singular causal perception. Other passages support this reading.

In the introduction, Michotte wrote (and I will add boxes for emphasis),

[2] "Koffka expressly mentions, in his textbook of psychology [*Principles of Gestalt Psychology*], that on Gestalt principles it is perfectly conceivable that one should have a specific impression of causality." [80, p.15,n.21]

> [Hume] has expressly asserted that in perceptual experience we have no direct impression of the influence exerted by one physical event on another. This assertion has been so widely accepted that it can still be regarded today as an almost universal assumption; [80, p.6], my box

The passage Michotte first cited for support is from the *Enquiry*, Sec.VII, pt.ii, reproduced here as written in Michotte (p.6). Italics in the following are in the original, but the box is added for reference:

> 'It appears that, in single instances of the operation of bodies we never can, by our utmost scrutiny, discover anything but one event following another.... So that, upon the whole, there appears not, throughout all nature, any one instance of connexion, which is conceivable by us. All events seem entirely loose and separate. One event follows another; but we never can observe any tye between them. They seem *conjoined*, but never *connected*.'

Hume argued not that there are no impressions of causality, but that *single* instances do not give such causal impressions. But although it sometimes appears that Michotte erroneously accused Hume of denying impressions of causality altogether, I think that by equating the two words boxed above, 'direct' and 'single', we can see that the debate centered on real differences in content, not misinterpretation.

Michotte thought Hume missed the causal impression because Hume relied too much on an analytical, mechanistic mode of thinking. In this analytical mode, Hume concentrated on what was really the case in the physical world, rather than on the psychological impression in the observer. Michotte allowed that Hume was correct about the real world, but argued that causality was a psychological phenomenon, and that it manifested itself readily under the appropriate combinations of mechanical movements, as he discovered in his collision experiments. As we saw in chapter 3, the singularists and process theorists denied that Hume had even represented the physical world correctly. This is the point made in the extended quotation from Humphreys in chapter 3 to the effect that Hume had performed logical operations that were not possible or even meaningful in reality.

Michotte took that mechanical causal impression as the basis of causality, and suggested that:

> If Hume had been able to carry out experiments such as ours, there is no doubt that he would ... probably have appealed in his explanation to the 'causal impression' rather than to habit and expectation. This causal impression, however, would have been for him ... nothing but an illusion of the senses.... It is probable that his philosophical position would not have been affected in the least.[80, p.256]

Well, certainly his antirationalist skeptical argument would remain unchanged. Would these experiments have convinced Hume to change his constructive account of causality? For the affirmative, we should note that Michotte, like Hume, advanced a psychological account of causality, and that Hume could have accepted an account which did not commit him to the reality of causation. For the negative, we note that, however loosely, Michotte's account is based on ideas involving force and dynamics, which Hume thought were unnecessary baggage, and that Hume need not have believed these experiments to demonstrate singular causation.

T.R. Miles, in his second critical essay at the end of his commentary on Michotte [81], examined the issue of Michotte and Hume. First, he underscored Michotte's separation of philosophy and psychology, and agreed that Hume's philosophical account remained unchallenged in the physical sciences even if it failed phenomenologically. However, there are at least a few issues where Michotte may be relevant.

First, if we accept a regularity account of causation, as Hume seems to imply we must, then we must accept that night causes day. Michotte's work provides some support for rejecting this claim, because it specifies additional kinds of perceptual organization necessary for the perception (or attribution) of causes. A dedicated Humean would reply that night does not cause day because although day always follows night, night also invariably follows day, and so although we have constant conjunction, there is no clear priority, so neither is the cause of the other.

Michotte adds to that Humean account a principled reason why the cause is the precedent condition rather than the resultant one. It also provides a way to distinguish causes from common effects. But on the phenomenological level, Miles wondered if there was really a disagreement. Michotte claimed that we "*actually see* causality," where a Humean would say we only saw successive movements, but the real debate here seems to be on the meaning of 'see.' According to Miles, "One can scarcely suppose that if a person says that all we really see is a moving knife and a static loaf, it follows that the situation looks different to him from what it does to the rest of us." [80, p.413] Hume did not denying the phenomenological impression of causality or necessity, but only the idea that it was perceived directly.

However, "Hume's formulations are so worded that at least they *seem* like phenomenology; he certainly *seems* to be denying that a causal impression in Michotte's sense ever occurs." [80, p.415] We have already spent time teasing apart this confusion. According to Miles, one reason it is so readily confused — many other psychologists do read Hume phenomenologically — may be that Hume himself would not have made the distinction between conceptual and phenomenological issues; Michotte, at least in translation, was not a model of clarity on this issue either.

Now, on the topic of the psychological origin of the idea of causality, Miles saw that there *was* a disagreement, and he considered Michotte to be on stronger ground. Michotte "has clearly shown that habit and expectation are not the crucial factors in giving rise to a causal impression," since first, participants see causes right off (if the conditions are right), and second, despite perfect correlations and repetition, participants seldom or never report causal impressions if the conditions are unfavorable for interpreting the display as a real mechanical impact [80, p.415]

That appears to be the standard reading of Michotte, but it cannot be maintained. The immediate perception of causes by experienced adults hardly conflicts with Hume. Hume's regularity account hinted at a far richer developmental sensitization to the regularities in the world around us, rather than some simple entrainment to local regularities during the course of an experiment. The few hundred trials shown to participants are a small fraction of the lifetime of experiences which observers presumably draw upon.

4.2.3 Putting it together

What Michotte's experiments might succeed in doing however is limiting Hume's suspicions: some regularities, no matter how prevalent, might not be perceivable as causal regularities. Michotte appears to have found some which at least are not so perceivable by adult observers. Other difficulties are known. Humans have a difficult time drawing causal connections over extended periods of time (food poisoning), or through multi-step chains. [47] Conversely, there are cases of single-trial learning, particularly where pain is involved, as Hume suspected in his example of the candle. These additional data points should help psychologists to construct a more accurate theory of causal perception.

Hume's argument against purely rational knowledge of cause and effect still stands. But Hume also claimed that we did not have *empirical* knowledge of the link, and here I think Michotte and others have made some inroads. In simple experiments Michotte demonstrated *constraints* on our perception of causality and at the same time prescribed conditions where it was almost impossible not to have the impression of causality. These conditions corresponded to objects behaving more or less in accord with the laws of mechanics, implying that people were sensitive to the observed behavior of objects.

In the richer experience of the world, there are far more constraints and far more information available to the observer than in Michotte's experiments, and more invariants to pick up on.

4.2.4 Michotte's theory

We have already seen Michotte's basic stance: regardless of the physical reality of causation, humans experience a causal impression, and this impression exhibits regularities. Hume, he claimed, would have seen this too if he had used a better set of experiments. Michotte thought that he had such a set, and tried to explain the perception of mechanical causality. He left qualitative causality, which includes long-term causes and less perfect correlations, for later researchers.

Michotte developed a whole family of theories over the years, revising them as new editions of his book came out. In an earlier version Michotte tried to sort the results on the basis of "polarisation," "approach," and "withdrawal," but later discarded these in favor of "phenomenal permanence" and "contrasting states." Rather than go through the evolution of these relatively *ad hoc* theories, it may be more convenient to look at surviving elements as discussed in more current literature.

Weir pointed out that "it is important to separate the possibility of inference making from the question of awareness of such processes." [114] She wrote a program to demonstrate how to accommodate many levels of inference in causal perception with Michotte's results. She also discounted some of Michotte's notions about what makes something the cause and something the effect.

> In one of these [Levelt's experiments], B is an elongated rectangle at midscreen. A travels to B and 'passes' in front of it, moving more slowly during the 'passing' phase and speeding up after it leaves B. The 'causal links' offered by participants are of the form 'resistance due to friction, viscosity, magnetic effects'. The important point is that the passive stationary object is now the agent of the 'braking effect'. This undermines Michotte's rather monolithic position. . . . [114]

Indeed, most of Michotte's interpretations of his results have not survived further testing. Neither have his claims for the universality of some causal impressions, or the lack thereof. It is thought that Michotte's reliance on "experienced" observers and subjective reports inflated the inter-subject agreement.

Subsequent researchers replicated Michotte's principal results in the launching and entraining experiments, but not to the same degree of assent that Michotte claimed. (See [21] for an overview.)

Other experimenters, such as Kruschke and Fragassi [69, 68], have also attempted to clarify and apply Michotte's thesis. Kruschke and Fragassi made computer displays of launching and delay events, and used these displays in a reaction-time task which was supposed to introduce the possibility of an objective measure. The results, however, were inconclusive, possibly due to difficulties with the displays.[3] What emerged was that observers have different reaction times for identifying tags between objects in a collision depending on whether the collision exhibited launching, coherent motion, delayed launching, nothing, or delayed nothing. Part of the difficulty with following Michotte is that he ignored the dynamics of real collisions in order to concentrate on a philosophical point.

Michotte conceded to Hume the logical point that "it would be perfectly possible to suppose without absurdity, even if it had never in fact been observed to happen, that B could set off on its own accord after being struck by A." Our belief that B moves only because A hit it is not a logical necessity. "It is rather a 'necessity' of fact, a purely phenomenal constraint, one which is not just a 'feeling of necessity' but one which is inherent in the event itself and is a character belonging to it." [80, p.263] Of course the observer must come to know what kind of behavior is inherent in which events, but Michotte thought that the feeling of necessity came first as a perceptual given, and repeated exposure helped observers to realize what was *not* necessary. For example, a child clawed by a cat might most naturally generalize that all cats claw, only differentiating through later experience. Humans and neural networks exhibit such tendencies to over-generalize, weighing first examples heavily, and applying the most common "rules" in all cases, until, at length, they learn otherwise.

For example, children learning to speak initially form the past tense of all verbs according to the patterns of the most commonly-used (and irregular) verbs, then realize the regular patterns used in most verbs (stem plus 'ed' in English), and over generalize *that* rule by applying it to all verbs, introducing new errors in verbs they had already learned ("I runned"). Neural networks that successfully model the rapid development of infant vocabulary also exhibit this tendency.

In the foreword to the English translation of Michotte's book, R.C. Oldfield summarized the contributions Michotte made to causal perception:

> We do not, according to Michotte, see one billiard ball cause another to move *either* because we intuitively apprehend a fact of nature, *or* because past experience leads us to see the event in this fashion, but because the spatio-temporal organisation is such that it directly unleashes this impression in us. Alter the relevant variables by a small but measurable amount and the impression disappears." [80]

Just what "directly unleashes" means has yet to be specified, although Michotte noted that the clearest impressions of causality were generated in displays that resembled physically possible events, and so might involve physical quantities:

> ... we can say that the causal impression is *the perception of the work of a mechanical force*, just as the impression of the movement of a car is the perception of its displacement in physical space. [80,

[3] (Fragassi, personal communication).

p.228]

> In short, in spite of the many contradictions mentioned in the course of this book, there is an impressive amount of agreement between the laws of mechanics and the properties of the causal impression. The correspondence between them is indeed so extensive that anyone not very familiar with the procedure involved in framing the physical concepts of inertia, energy, conservation of energy, etc., might think that these concepts are simply derived from the data of immediate experience.... [80, p.263]

Michotte had a good intuition here, but no precision. Is he talking about force, energy, momentum, or inertia, or some combination of all of them? In our common conception, causes are associated with forces: both causes and forces make things happen. But if Salmon and Dowe are right, then causal interactions involve the exchange of *conserved* quantities, not force. The quantities are related, but here the Dowe-Salmon formulation offers us the precision which Michotte did not pursue, and provides some philosophical grounding for finding causes in particular quantities.

As Johansson has noted, Michotte took psychology out of the old paradigm of inferences from static images, and into the world of dynamic events extended in time. Energy is a property not of objects in cinematic freeze-frame, but of objects in motion. Michotte showed that participants extract some sort of impression from dynamic displays, but his experiments were artificial animations done without regard to conservation laws. It remains to be seen if observers can perceive physical dynamics, and if we can distinguish perception of force from perception of exchange of energy.

We know that people do not have good conceptual distinctions between the concepts of force, energy, momentum, and the like. Many experiments in naïve physics tell us this, and so does the history of 17th-century science from Descartes to Newton. On the other hand, people manage pretty well in the world, so at some level they have to be noticing relevant relationships and information which nevertheless never make it to representational, conceptual knowledge. So what is it that people are perceiving?

The current consensus is that humans are very bad at perceiving acceleration and other second-order (second derivative) properties, and this gives reason to think that force or a force-analogue is not directly perceptible. However, energy and momentum derive from velocity, and so are first-order quantities, which can be specified directly in the kinematics. As we shall see, relative mass is also specified in the kinematics.

In other words, the data may be richer than traditionally supposed in Cartesian-inspired perceptual psychology, rich enough to allow human observers to solve the problems of perception. Of course, that this problem in inverse dynamics which we call perception[7] can be solved at all means there are constraints on allowable solutions, and these constraints are likely to be the peculiar physical constants of our environment, such as $g = 9.8m/s$, which learned through experience. These are called "uniquity conditions."

There is no reason to believe humans would have any perception-action difficulties if raised on the moon, but of course we do expect the transitions between regions with different laws to require some adaptation, not unlike that experienced by participants adapting to wearing displacement prisms. There are good evolutionary and developmental reasons for believing that there is a place for Hume's repeated experience, since it leads to a much more robust system, and to believe that the correct place is in the detection and use of uniquity conditions to facilitate solutions to the dynamical problems of perception. If human abilities *can* be expressed in this language, then the stage is set for merging causal perception with automated causal induction.

Michotte's experiments suggest that adult participants have become attuned to the regularities in our world, that they are now biased towards certain mechanical relations to the exclusion of other regularities. Michotte supposed that the perceptual system was hardwired to pick up on these kinds of relations. One mechanism that other have suggested is retinal feed-forward processing. But it is not clear just how "hard" that wiring is. In adults it may no longer be plastic, but if something like Edelman's *Neural Darwinism* thesis is correct, might the retinal wiring have been selected because of the laws of the world we grew up in? Even if all of Michotte's studies were upheld, the question of the hard limits of causal perception is still unresolved. If we are to resolve them at all we need to incorporate the laws of motion into our event-perception studies.

4.3 Innate and Learned

There is no real prospect for resolving the issue of relative contributions of innate versus learned aspects of event and causal perception. The largest problem is that there is no longer a useful border between the two ideas; this is not the place to get into the issue. I mention only in passing two areas of research which have a bearing on where the border, fuzzy as it is, falls. The first is the study of causal perception in infants. The obvious way to decide between innate (Michotte-style) and learned (Hume-style) accounts is to trace causal perception back towards (or before) birth, and see if there is ever something that looks like an onset, separable from other prerequisites. Obvious but beset with theoretical and practical problems. The other is to look at the neurobiology of event perception for some analogue of or explanation for a percept of causal interaction.

4.3.1 Infant Causal Perception

Infant perception studies are notoriously problematic. If we do believe them, then it appears that children preferentially use a notion of "generative transmission" over "other rules such as covariance, spatial contiguity, temporal contiguity, and regularity of succession," [70], indicating an adherence to some kind of "causal power" conception, which may be in line with Dowe and Salmon. In one example,

> children of 2-4 years would identify an already operating fan as the cause of a candle blowing out if a gap in a windshield was oriented towards it just as a second fan was switched on. The fan which was capable of transmitting energy to the effect was selected as a cause, in preference to a second fan which was not capable but whose onset of action was temporally contiguous with the effect. [70]

Other work indicates that concepts of physical causality may develop over time, and that younger infants may be less able to ignore contiguity and pay attention to processes [88, 71]. Even if we do not believe the results of these studies, there are a couple of issues which need to be addressed regarding the interpretation of Hume and Michotte, and the possibility of testing between them.

Leslie and Keeble [71] began their article "Do six-month-old infants perceive causality?" with a paragraph summary of Hume in which they wrote,

> If Hume had ever considered infancy, he would no doubt have thought that infants, lacking any substantial experience of the world, would only be able to sense the spatial and temporal arrangement of events, and have little or no knowledge of causality.

Hume did argue that we would learn causes through prolonged exposure to a constant conjunction of objects, and that our knowledge of causes was nothing more than this mental force of belief engendered by such prolonged exposure. But Hume did not specify what "prolonged" meant. Leslie and Keeble, along with some other psychologists, have assumed that six months is not long enough, and so "refute" Hume's position by showing that six-month-olds already have a notion of causality.

Instead, Leslie and others have shown that infants develop an increasingly refined sense of causality between 6 and 10 months, a result which could just as easily be read as support for Hume's position that we develop our notions of cause and effect through our participation in the world. Perhaps causality at 6 months is surprising, and perhaps Hume would not have expected it, but 6 months is rather a long time, and the infant experiences much of the world in that time.

Leslie and Keeble also wrote,

> there are two traditional but opposing hypotheses about the nature of this [internal structure of movement] relationship as naively perceived. On the one hand, Hume would argue that infants will perceive two independent aspects of the event—the spatial contact and the temporal succession of the movements. Against this Michotte (in company with Gibsonians) asserts that a causal relation will be registered directly. [71]

Since Hume argued that the mind adds a strong causal perception, it is hard to test the difference between these two theories. How can one separate direct perception from unconscious but not quite-as-direct perception? The real question is when the transition to causal perception takes place. Hume has already hypothesized that it does take place, but he has not said when. Michotte at first appears to believe some causal perception is built in from the beginning. This also is too simple, since infants must develop the ability to focus upon, track, and perceive objects, all of which are prerequisites for visual perception of causation.

Perhaps a more interesting question is whether there is any time during development when there is meaningful perception of objects without perception of causality. If not, then there must be limitations to Hume's theory, which is to say some more innate component to the perception of causality.

In contrast to Leslie and Oakes, Wickelgren and Bingham [117] tested eight-month-old infants on natural and unnatural events, paying strict attention to the dynamic analysis of trajectory forms, and found no difference in looking time for natural (forward) or unnatural (reversed) events, even those which are normally asymmetric with respect to time. To the extent this more detailed analysis becomes replicated, there is some evidence for a more Humean thesis which separates object perception from event recognition, or at least the distinction between natural and unnatural events. Having covered both sides of the debate it is probably wisest to resolve that the issue is unresolved, given the uncertainty of looking-time assays and the difficulties in ascertaining just *what* the infants are responding too.

4.3.2 Neurobiology

I just want very briefly to call attention to work done on retinal processing of motion and in particular on motion priming. Visual edge-detection (and enhancement) circuits exist at the retinal level and are fairly well understood. More recently, researchers have discovered that very simple retinal circuitry can explain the important but previously mysterious ability of the visual system to keep up with a moving stimulus which is tracking across the visual field faster than the known delays in the conduction and processing of visual signals from eye to brain. [40, 6] Berry *et al* wrote,

> A flash of light evokes neural activity in the brain with a delay of 30–100 milliseconds.... A moving object can cover a considerable distance in this time, and should therefore be seen noticeably behind its actual location. [But it is not.] ...A moving bar elicits a moving wave of spiking activity in the population of retinal ganglion cells. Rather than lagging behind thte visual image, the population activity travels near the leading edge of the moving bar. This response is observed over a wide range of speeds and apparently compensates for the visual response latency. We show how this anticipation follows from known mechanisms of retinal processing. [6]

The "gain control" mechanism works by inhibiting ganglion cell firing after initial onset, thus effectively shifting the peak forward in time so that the peak coincides with or precedes the leading edge of the bar. Such a neural mechanism amounts to a motion "edge detector," enhancing the perception of the moment of passage of a leading edge, *only* if that edge is moving, and in proportion to the velocity of its retinal motion.

The phenomenal effect which is thought to derive from this has already been demonstrated, and is described in Gegenfurter's summary [40]. If a rectangular bar is rotated around a central point, and observers are asked to report its position at the time when a stationary bar is flashed alongside it, observers report the moving bar to be in advance of its actual location at the time of the flash.

4.4 Naïve Physics and Event Perception

Naïve[4] physics is the research tradition in experimental psychology which studies subjects' naïve beliefs about simple mechanics. Although most famous in the context of education, the field began in the Artificial Intelligence community. Intelligent agents must be able to solve the *frame problem*, which is the problem of the scope or limit of actions. If block B rests on block A, and I pick up block A, block B will come along, or fall off. Humans solve this problem without difficulty, and usually avoid grabbing the middle plate in the stack. Computers have a difficult time. Either a great deal of explicit knowledge has to be built in, or they have to have some notion of causal structure [66]. In short, they need a naïve physics.

In his revised manifesto, Hayes advocated the formalization of tacit physics, and described his program as formulating the grammar of causality, slowly axiomatizing observers' internal beliefs. [49] That the project is feasible at all is supported by McCloskey's finding that this naïve physics seems to be largely

[4] Meaning "atheoretical" or "inexperienced."

consistent across individuals, even though it is markedly inconsistent with Newtonian physics. [78] Whatever its motivation, the core of the field is the experimental analysis of subjects' own theories of motion, often elucidated through "what if" sketches of possible physical situations. Participants are usually asked to complete an object trajectory, or label the forces on an object at a particular time, or some similar task.

The original studies were almost exclusively in the descriptive mode, which compromised their claim to discover the native theory humans actually use to get around the world, since they are not in the least perceptual experiments. The studies are still very relevant for physics education however, since they show how difficult it is to get humans to change their explicit conceptualizations of how the world works, and can suggest ways to make the teaching of physics more successful in that endeavor.

More recent studies have borrowed the approaches of event-perception research and moved towards the perception-action end of the experimental spectrum. Consequently, there are new conceptualizations of humans' naïve beliefs about the world. After a brief criticism of the descriptive approach, I will turn to work that lives closer to the event perception tradition. I bring in the work on naïve physics because even though the event perceptionists distance themselves from the field on account of methodology, in fact much of the work on the perception of dynamics and naturalness is a kind of naïve physics of tacit, rather than explicit knowledge.

4.4.1 Criticism of early interpretations

On the descriptive end we find two studies in the classical naïve physics tradition. McCloskey [78] and Clement [18] gave their subjects written scenarios, and asked the subjects to describe what would happen, and explain why. For example, shown an overhead picture of a ball whirled on a string, subjects were asked to draw the path of the ball if the string were cut. Even after a year of physics, 30% drew curved instead of tangential trajectories. When asked about balls emerging from flat spiral tubes onto a flat tabletop, 51% drew curved trajectories. Balls dropped from moving planes apparently go straight down (36%), down in a diagonal (13%), or backwards (11%). Understandably, this appalled physics teachers, although one can hardly believe it came as a surprise. More significantly for us, psychologists took this and other examples as evidence for what our own internal naïve theories of "physics" must really be like. In some cases [18] researchers explicitly proposed that subjects' knowledge of physics recapitulated the history of science, and identified the naïve theories with pre-Galilean impulse theory.[5]

The problem is that while this kind of experimentation is good at showing how hard it is to teach physics, it conflates analysis and action. If by "naïve theories" we mean those implicit theories used by humans to navigate the world, then we are in the domain of perception-action, not descriptive analysis. The results of McCloskey's study do not square with the perceptual experiments on real dynamics and event recognition. The difference is between riding a bicycle and analyzing what is involved. For example, while coasting down a hill at any decent speed, one begins a left turn by pushing on the *left* handlebar, not the right one, although few people are aware of this. There is clearly a divorce between tacit and explicit knowledge.

Nevertheless, Proffitt, Kaiser, and Whelan [92] made the mistake of expecting cyclists to have better *explicit* knowledge of wheel dynamics and gyroscopy,

[5]The question of *which* mediæval impulse theory was more rarely discussed.

presumably just because they are around spinning objects more often than the rest of us. When the cyclists showed no real difference[6] the authors concluded that experience did not help. But of course cyclists spend no more time rolling wheels down inclined planes than any of us. On the other hand, arguably, high-school physics teachers do spend more time at such things, but they too did no better than the control group.

What the article established is that tacit knowledge does not imply explicit knowledge. Even if subjects tacitly knew that it would take more effort to spin up a rim wheel than a uniform wheel, extrapolating from that knowledge to a prediction about which will descend faster on a ramp is reasoning, not perception.

In a summary article, Proffitt and Gilden [91] reviewed an experiment where they repeated McCloskey's earlier studies, but with the addition of computer simulations. As McCloskey found, subjects were poor performers on the descriptive task, but they quickly recognized as anomalous the behavior of simulated pendulums which were made to act in accordance with subjects' predictions. Proffitt and Gilden thought the difference was that animated displays separate the different contexts and dimensions of motion into different times, so subjects can use their unidimensional heuristics.

The cue and heuristic-inference approach drives Proffitt and Gilden's research program. This inferential tradition goes back to Helmholtz and early psychophysics and stands in opposition to the direct-perception line running from Mach through Gibson and Johansson. Although I suspect the empirical predictions of the two research programs can be made to agree, the metaphysical opposition results in different attitudes towards perception. Ecological theorists start first with a complete specification of the information available in the *actual, rich* environment, and presume that subjects can make use of most of this information. In contrast, inference-based approaches start with analyses of simple controlled cases to learn what subjects can and can not perceive, and concentrate on failures and biases to help find out what rules or heuristics subjects are actually using in the pure cases.

Ecological theorists argue that simplified cases are just unrealistic, and can not serve as base cases. For instance, since Descartes if not earlier, researchers have wondered how we take a 2-dimensional projection on the retina and turn that into a 3-dimensional conception of the world. A natural first step in analysis is to simplify the problem by fixing the eye and investigating what information it can and can not use. However, it turns out that the fixed eye not only can not solve the problem, but in humans a truly fixed eye saturates the receptors in a couple of seconds and can no longer see anything at all. This base case was fatally flawed: for human visual perception the moving eye is an essential part of any visual perception let alone space perception. Once the moving eye is considered, several problems in the extraction of 3-dimensional space are solved.

This debate about the role of rich data versus simple cases is reflected in the debate between causal modelers and their opponents. Causal modelers think that the full data are rich enough to provide constraints and allow us to perceive or extract causes from covariations. Opponents tend to note that we must fail in principle on the base case, and think therefore that more data only makes the situation worse. This methodological standoff has manifested itself in the event perception literature as well.

---

[6] They did do better on explicit questions about bicycles, but the results do not seem generalizable or even very meaningful, and they can be explained largely by the fact that the cyclist group had on average at least one more year of college and two more physics courses than the "naïve" control group.

4.5 Real dynamics

There are two main kinds of dynamic simulations in the event-perception literature: two-dimensional two-body collisions, and gravitational motion. The collision studies are the more direct descendants of Michotte's animations, only with real physics. The key result for this paper is that under the proper conditions, subjects can judge the relative masses (a dynamic quantity) of the two colliding balls, given only the motion displays. Exactly how subjects do this is a matter of some debate. Sverker Runeson [96, 95] argues that the Kinematics Specify the Dynamics, which argument is known as the KSD theory. By this he means not only that the conservation laws imply that the relative mass is uniquely specified given the relative change in velocities $-\Delta v_1/\Delta v_2 = m_2/m_1$, but that this information is used by observers. His opponents [42, 91, 92, 41] counter that humans are sub-optimal in ways that indicates we use simple heuristics instead of perceiving the dynamics. The debate continues, but Runeson's opponents have at least retreated far enough to allow that with experience, human performance can more closely approach the ideal, even if observers start out relatively poorly.

The two sides have very different ideas of what constitutes a proper experiment. Proffitt and Gilden attempt to isolate specific factors in controlled experiments, Runeson criticizes them for failing to be "ecologically valid" and performs broader experiments exploring a larger range of parameter space. This debate is on the same axes which define the laboratory-versus-field debates in biology. The closer we approach to ecologically valid conditions, the more applicable and the less comprehensible our results. So far collision studies show that under a wide range of initial conditions, subjects extract reasonably accurate perceptions of the relative mass of the colliding balls, although they can be fooled. We will look at the debate in more detail shortly.

Because there are many conditions where the kinematics of two-body collisions specify mass differences, collisions are not an ideal place to begin an investigation of the perception of energy and momentum; manipulations of energy or momentum of the objects involved confound specifications of the mass of those objects. Switching to displays of one object falling freely under gravity removes this confound, because the trajectory of fall is independent of the mass of the falling object. Later we will want to revisit collision studies.

Interest in free-fall trajectories in event perception is motivated by the ability of gravitational trajectories to specify distance. Distance or depth perception is a major problem in visual optics, since human binocular parallax is not adequate to explain human performance. Indeed, as many authors point out, humans get good depth information monocularly, so there has to be more information. A touchstone paper in this literature is Saxberg's analysis of the information specified in gravitational motion [100]. In short, the trajectory scales with, and so specifies, distance. Saxberg conducted a study to try to isolate this variable and see if observers used it [101]. However, in isolating his variable, he kept image size constant, so that as the ball approached, its image did not expand, resulting in conflicting information [79].

Muchisky and Bingham simulated bounces of various balls at various distances, chosen so that the image size of the ball was always constant. Subjects shown displays of various balls at various distances, of the *same optical size* could tell how far away they were by how they fell [82], indicating that observers were using the dynamical information. Another study, by Hecht, Kaiser, and Banks[51], suggested inconclusively that subjects might use only average velocity instead of acceleration, but supported the conclusion that observers could get relative distance information.

A recent study by McConnell, Muchisky, and Bingham[79] attempted to isolate the role of trajectory form more directly. Trajectory form in a resisting medium depends on distance and object density. The investigators found first, as in previous studies, that subjects were good at judging relative distance. Second, they found that given feedback, subjects calibrated quickly, and generalized that calibration. Third, they found that if object density were manipulated, subjects interpreted that as changes in the object's distance. In this study, both simulations and real objects (in patch-light conditions) were used.

Other events also provide information on gravitational dynamics, and Pittenger[90, 89] has done several experiments on pendulum motion which demonstrate that given real stimuli, observers are sensitive to the dynamics. The equation of motion for an ideal simple pendulum is

$$p = 2\pi k(L/g)^{1/2} \tag{4.1}$$

where p is the period, L is the length, g is acceleration due to gravity, and k is an angle-dependent coefficient which is essentially constant for angles below 30 degrees.

In his first study, Pittenger found that subjects could judge the relative lengths of pendulums in motion even when the bottom of the rod was obscured [89] (see figure 4.2). This implies that they are sensitive to gravitationally-specified natural motion, and have some intuitive knowledge of the physical law (*e.g.* "longer pendulums are slower"), although without absolute scaling.

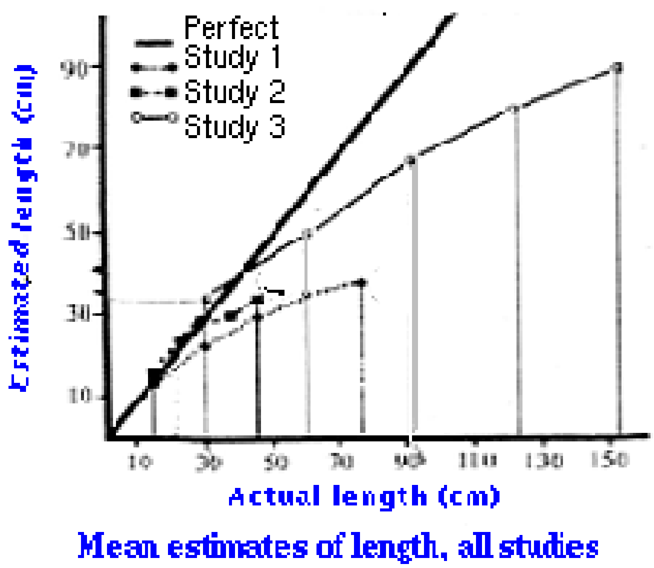


Figure 4.2: Estimates of pendulum length versus actual length. Subjects could only see the top part of the string, and so had to rely on period only. Subjects can judge relative length, but exhibit systematic underestimation with increasing length.

In a second experiment [90], subjects asked to judge the naturalness of a pendulum's swing could do so reasonably well. The experimenters made the pendulum swing faster or slower than normal by driving it with a second pendulum hidden from the subject (see figure 4.3). A free-response task let the experimenters explore the phenomenology. Subjects did not spontaneously report forces in the anomalous displays, but they did spontaneously describe the natural motion as "free of force," which implies that people do not see gravity as a force at all, but a natural part of object behavior.[7] On the other hand, occasional experimenter errors resulting in out-of-place accelerations were detected immediately.

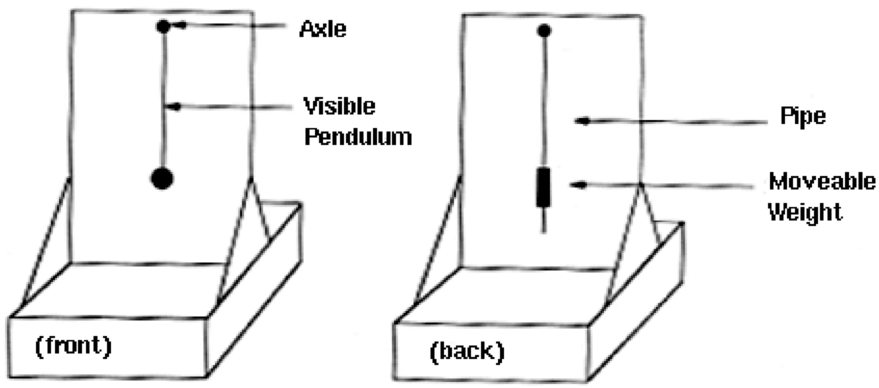


Figure 4.3: The experimental setup used by Pittenger in his 1990 study. The front of the apparatus, shown on the left, consists of a very light pendulum which is driven by the more massive adjustable pendulum in the back of the device, shown on the right. Subjects were asked to judge the naturalness of the motion of the visible pendulum.

4.5.1 A debate over collisions

Over the past two decades Sverker Runeson has carried on a debate with Proffitt and Gilden. Proffitt and Gilden have argued, among other things, that humans have an innate inability to deal with gyroscopes and other rotating or other multidimensional systems such as 2-body collisions, rolling wheels, balances, and fluid displacement problems [41, 91, 92, 42]. Runeson [96, 95] has countered with several collision studies which show that people can use the available information, and with training, use more of it.

This debate began with Runeson's 1977 dissertation, and his subsequent work on the visual perception of mass and other dynamic quantities. Runeson's work follows Johansson's theory of perceptual vector analysis [118, 61, 60]. Runeson and Vedeler [96] summarized the Johansson work saying, "the finding common to these studies is that perception is fundamentally based on velocity vector components of the motions of the elements in the display," and support this theory by showing how change in collision-axis velocity predicted their subjects' responses better than absolute exit velocity in their third experiment.

Proffitt and Gilden originally appear to have made impossibility claims about the nature of human perceptual processing, including the impossibility of directly perceiving abstract constructions like the collision and sweep axes of

[7]This, incidentally, may be a plus for Salmon, who regards gravitational fields as geometrical structures of spacetime, and free-fall in such fields as geodesic motion, in concert with relativity theory. Theorists in the ecological perception world would state this as attuning to invariants in the environment.

motion. Runeson and Johansson's experiments together have shown that at least some abstract axes are perceived, in particular, that common motion is extracted (and ignored) where possible. Runeson complained that Proffitt and Gilden used overly constrained conditions in their collision studies, and that therefore their conclusions do not apply in the general case. Runeson thinks the data are still consistent with his KSD (Kinetic Specification of Dynamics) theory than with their heuristic approach [95]. Runeson believes that observers can directly perceive the dynamics of the situation by noticing invariants in the rich kinematics. It will be helpful to trace the debate in some detail.

Basic introduction

In a general two-dimensional collision, the mass ratio is specified by the *change* in velocities, implying that pre-collision velocities are important. Runeson analyzed collisions in terms of sweep-axis and collision-axis, but warned (p.1262) that the visual system didn't need to use that particular decomposition, although that's the one he seems to have in mind outside that caveat. Runeson spends several pages laying out the supposed clash between cue-heuristic theory and KSD. In fact it is a description of the conflict between the Gilden and Proffitt and the Runeson and Vedeler theory. In short: Gilden and Proffitt argued that humans cannot use multidimensional information, or information relying on memory. Runeson and Vedeler disagreed. Runeson was explicitly Gibsonian in his argumentation, seeking "liberation from elementaristic ontology"

Gilden and Proffitt 1989

We enter the debate in 1989 with Gilden and Proffitt's attack on Runeson. [42] In an article titled "Understanding Collision Dynamics,"[8] they used simple two-dimensional collisions to establish deficiencies in Runeson's KSD theory. Gilden and Proffitt claimed that humans just use two heuristics to determine mass ratio in collisions: the velocity heuristic and the angle heuristic.

Velocity Heuristic (VH): ball exiting with much greater velocity ($> 2\times$) is lighter[9]

Angle Heuristic (AH): if incoming ball *reflects back*, it is lighter.[10]

By showing that when the two heuristics AH and VH conflict (incoming ball rebounds, second ball moves away with twice the $|v|$ of the rebounding ball), people became much less accurate, Gilden and Proffitt supported their theory that people use something like these two heuristics instead of actual mass ratio.

Gilden and Proffitt claimed as a result of their first experiment that pre-collision velocities are "formally superfluous" to the determination of mass ratio, at least *when M_2 is initially at rest*, as it was here. That is in itself an odd restriction for a paper which justified its experiments by claiming, "A proper analysis of the structure of mass ratio judgments cannot be accomplished unless [observers] are allowed to view collisions in their most general form." Runeson and Vedeler would reply, quite rightly, that the Gilden and Proffitt result was explained by the fact that there was very little pre-collision information *difference*. In the occluded condition, a bar was on-screen ostensibly to occlude the ball, but of course it thereby showed the path the ball was taking, which is

[8] A companion to "Understanding Natural Dynamics," cited previously, where they reviewed and extended the classical naïve physics program.

[9] "The ball with the greatest velocity in the post-collision epoch appears to be lighter."

[10] "The ball that scatters at the greatest angle appears to be lighter."

pre-collision information. In addition, the ball appeared at collision, potentially giving the collision angle. Finally, it is not clear that there is no difference. If their figure 3 is re-plotted as a discriminability function (Y axis is "% M_2 is chosen as heavier"), then the unoccluded condition is closer to a zero intercept, representing veridical judgment.

In their second experiment they found support for heuristics. There were two tasks: a binary decision and crude magnitude estimate. Observers were presented with 3 collisions in each of the 4 possible mixed salience conditions for the 2 heuristics. Observers did best (83% correct) when only 1 heuristic was salient, and worst (62% correct) when both AH and VH were salient. In the latter case, the rapid exit velocity of the heavier "ball" confused observers into thinking it was lighter. When neither heuristic was salient, observers did medium well (77% correct).

Comparison with Todd & Warren 1982

Gilden and Proffitt compared their results with a study by Todd and Warren in 1982 [110], who showed that in *one-dimensional* head-on collisions, observers were virtually perfect when $m_1 > m_2$ but around $m_1 = m_2$ were very poor. The Todd and Warren results, as summarized by Gilden and Proffitt are shown in in table 4.1. As Gilden and Proffitt indicated, their two-heuristic model was better able to explain the results than the one-heuristic model presented by Todd and Warren. Gilden and Proffitt took this retrodiction as additional confirmation of their model.

m_1/m_2	v_2/v_1	% correct
> 1		almost perfect
0.67	-4	48
0.50	-2	75
0.33	-1	90

Table 4.1: Todd and Warren 1982, per Gilden and Proffitt.

However, Todd and Warren's results were more detailed than shown in table 4.1. In fact, m_2 could be either stationary in the center of the screen, or moving towards m_1. I reconstruct the Gilden and Proffitt table using the more complete Todd and Warren data in figure 4.2 The results for the expanded data are graphed in accordance with Gilden and Proffitt's presentation order in figure 4.4.

m_1/m_2	% corr. (moving)	% corr. (stationary)
0.33	99	90
0.50	95	75
0.67	95	48
0.80	89	38
1.25	90	93
1.50	98	96
2.00	96	95
3.00	98	96

Table 4.2: Todd and Warren 1982, Experiment 2, p.330, $e = 0.9$ condition

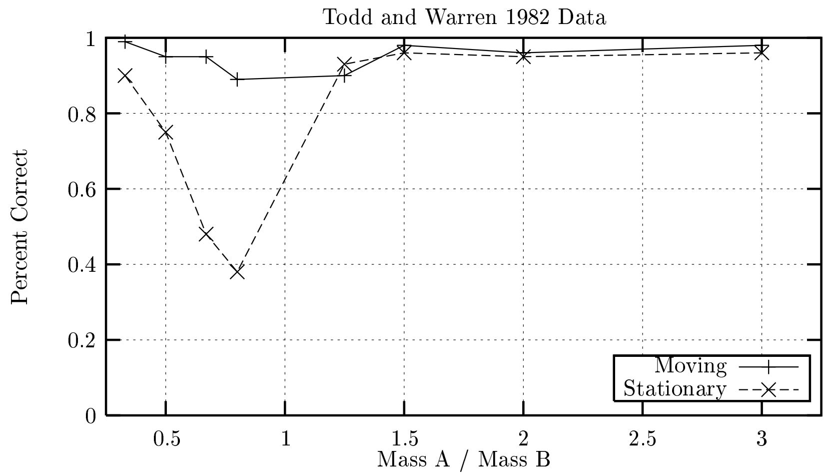


Figure 4.4: Todd and Warren as Gilden and Proffitt (1989) saw it

Gilden and Proffitt are perhaps justified in comparing their experiment with the most relevant version of Todd and Warren's (column 3 in table 4.2), but it is worth noting that observer performance improved markedly in the more general condition where both blocks were moving in the pre-collision epoch. In fact, on the face of it, this supports the claim that pre-collision information *is* useful.

Runeson & Vedeler 1993

The main point of Runeson and Vedeler (1993) was to poke holes in Gilden and Proffitt. [96] Gilden and Proffitt are charged with irrelevance because the conditions were too restricted and because they used threshold measurements. Three new experiments with richer variability favored KSD over two-heuristic models. Runeson and Vedeler did not point out the reinterpretation of the Todd results. However, in making methodological criticisms of the way Gilden and Proffitt presented their data, they argued that Gilden and Proffitt used the wrong kind of graph to understand the results of their own experiments. According to Runeson and Vedeler, Gilden and Proffitt's data should not have been graphed as "% correct," but "% chose object x." The motivation is that Runeson and Vedeler argued that the observers are really being asked to do a discrimination task between masses, for which the second kind of graph is more appropriate.

Runeson and Vedeler make their points with regards to the Gilden and Proffitt data. Taking their hint, we can change the coordinates on figure 4.4. If, instead, we graph the Todd and Warren data on discriminability axes, the odd dip in performance in the middle of figures 4.4 disappears, and we get a classic logistic (ogive, S-shaped) curve, as shown in figure 4.5.

In figure 4.5, it is easier to note that, unsurprisingly, observers are worst at discriminating mass when $m_1 \approx m_2$. What *is* interesting out of all that is that it yields an obvious interpretation of the relatively poor performance in the stationary condition: there is a bias towards the impinging mass. In fact, this bias is a robust effect throughout the debate, and is referred to as the offset in the point of subjective equality (PSE).

It's important to realize that Runeson and Vedeler were doing much more general collisions than Gilden and Proffitt. Runeson and Vedeler had two balls

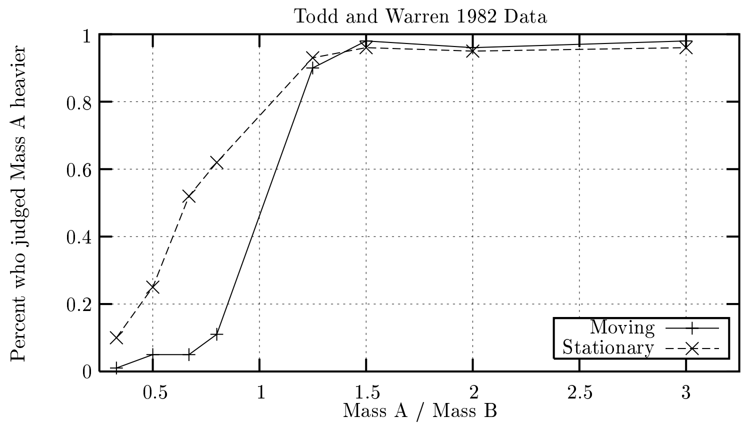


Figure 4.5: Todd and Warren presented as discriminability data

coming at different speeds and angle from different parts of the screen, colliding, and flying off. Therefore, they found clear advantages when they did not obscure the pre-collision epoch. More importantly, in order to apply Gilden and Proffitt's heuristic model, they had to reinterpret VH as a *win-velocity difference*: the difference in the collision-axis velocity component, subtracting the "loser" from the "winner." The winner is the ball which successfully invaded the other's half of the screen.

Endgame

In a reply [41], Gilden conceded observer ability to use pre-collision velocities at least sometimes, but overall defended heuristics at a more general level fairly convincingly, effectively putting Runeson in check by using his own data. Gilden used the more general model on Runeson and Vedeler and claimed a good fit. In a reply [95], Runeson charged that the model had too many but his detailed critique was essentially nitpicky, and he conceded that the cue-heuristic model "appears to be consistent with the direction and overall magnitudes of the PSE offsets".

In what I can only regard as a desperate defense, Runeson [95] sacrificed two of his own experiments from Runeson and Vedeler (1993) in order to rebuff the attacks in Gilden (1994) and save KSD. Runeson reanalyzed the data from the most detailed (third) experiment of Runeson and Vedeler (1993)—the one not analyzed by Gilden—and found that observers did better than a generalized two-heuristic model would predict, that the errors were not of the type that a two-heuristic model would predict, and that a model based on collision-axis velocity components fit the data better. However, he had to dismiss previous studies (including his own) because they were based on "suboptimal observer performance," or "because of confounding and lack of representative variation in parameters."

The third experiment was the most detailed to date in the literature, and interestingly, the one with the least PSE bias (21%). After performing some ANOVAs and regressions, Runeson concluded that mass ratio (the KSD quantity of choice) was more informative than scatter-angle or exit-speed ratio (the supposed "cues"). However, we don't get a comparison with the straight combination of the two cues. What Runeson does give is a prediction of the data given his best guess at a generalized cue-heuristic theory. The result was that observers seem to have lower PSE biases than Proffitt and Gilden's cues would predict, which appears to be somewhat of a vindication of KSD.

Runeson got such a result by an experiment designed to confound the exit-speed ratio. Runeson and Vedeler varied sweep-axis velocities, which are unchanged by the collision. Thus, they could increase the exit velocities as they desired. The VH would require those faster-exiting objects to appear lighter, even in the face of strong scatter-angle differences to the contrary. By using this kind of design Runeson and Vedeler were able to find a realm of collision where the cue theory was at a disadvantage. In addition, experiment 3 used experienced observers.

Stalemate?

The constant revisions and methodological sparring in the debate led Hecht to question whether, in fact, anything was being debated at all. [50] In a 1996 literature review he asked whether the competing research programs had become immunized concepts or merely non-statements. As far as he could tell, there was no longer any essential difference between KSD and heuristics.

In an as-yet unpublished study done in 1994, Hecht and Proffitt had compared the performance of billiards players to their conceptual understanding of the game. [52] Both novice and expert players were tested on perceptual and conceptual tasks. The result was that the novice data was well-explained by a heuristic approach, and that even experts had incorrect conceptual knowledge. However, expert performance on judgment of complex events nearly rivals their skill in producing such events, so heuristics are not enough. In short, experienced persons can come to know multidimensional info, contrary the Proffitt and Gilden's heuristic approach, but do not generalize to the extent that KSD would imply.

What has happened is that KSD theorist have moved explicitly into investigations of the possibilities of human performance. Using experienced observers, in data-rich environments, they ask what observers *can* learn, given the chance. Cue-theorists, on the other hand, are more interested in decisions and judgments made in non-ideal cases. Both sides think their emphasis more closely reflects real-world perception.

4.6 Conclusion

Like time and space, causation is not a primitive sensation, and so we have to explain what is meant by the perception of causality. In the middle part of the twentieth century, French perceptionist Albert Michotte undertook the direct investigation of causal perception. In a very straightforward manner, he showed observers animated displays of collisions, and asked them to describe how "causal" the interaction appeared, or to free respond. However, "causal" is not a well-defined term. One can only fear what would have happened had he run a philosophy graduate student on that task. It is clear that Michotte's numerous experiments were getting at *something*, but it is hard to tell just what. The whole "gestalt" methodology under which he worked has come under strong criticism for exactly this kind of ambiguity. Broadly speaking, however, his subjects were more favorable to collisions that approximated possible real-world collisions, at least so far as any of Michotte's collisions were representative of real-world collisions.

Indeed, one of the problems is that Michotte's background theory and design did not take real dynamics into account. Perceptionists in the latter half of the century have begun to work with more accurate simulations of real dynamics, in collisions, gravitational motion, and other sorts of physical interactions. The field in which this work takes place is known as "event perception." The goals of this field are to answer standard psychological questions of recognition, sensitivity, and categorization, about events as opposed to objects. Events are extended in time, although very little is known about natural segmenting of events. Runeson [94] drew distinction between events, which are relatively brief, and processes, which are indefinitely extended. Interestingly, his examples seem to coincide perfectly with Salmon and Dowe's distinction between processes and interactions. For my purposes I prefer to let the term "event" remain broad enough to cover events of all durations, and to distinguish among events in terms of processes and interactions, which are more principled distinctions. However, the merit of bringing in psychology is to help make distinctions in the world which accord with the natural segmentations of our perceptual system. The motion perception system is a complex oscillator embedded in time, and its circuits have default rates which likely determine a qualitative difference between interactions of shorter or longer durations.

One thing we have seen in some of the event perception studies, most particularly those of Pittenger [89, 90], is that while asking about perceptions of "causality" straight off are suspect, observers *are* able to perform "naturalness" judgments. With well-defined natural events, such as real pendulum motions, observers were very sensitive to deviations from natural swings. Other experiments using computer displays have shown that observers can tell by the gravitational dynamics the proper distance of objects in the animation. [51, 79, 82], and conversely, can adjust gravity to the proper value based on the distance a fountain is placed in a three-dimensional backdrop. [106] For these reasons I have argued that the correct way to study causal perception is through event perception. In chapter 5, I will present two new experiments which investigate the perception of causality using observers' naturalness judgments on events where certain conservation laws are violated to make the events unnatural in such a way that, on the CQ theory, they specify causal interactions that not only are not apparent, but in fact whose dynamics correspond to nothing which could be inferred on the basis of expreience.

Chapter 5

Two New Experiments

> Causation may not be in the real world or in the equations, but it definitely is in our thinking. Hayduk

The existing experimental literature has explored the perception of causation (Michotte) and the perception of dynamic quantities (Runeson). On the one hand, Michotte and others who have repeated his experiments have not pressed the issue of exactly what properties were perceived. In part this is because they have not taken a simulation approach. On the other hand, authors in the latter tradition ask what properties are perceived, but do not apply their work directly to the problem of *causal* perception. There is good reason for this: the problem is vague, and even if it is resolved, it only accounts for part of the vast domain of causality.

The experiments presented in this chapter were designed to evaluate, indirectly, whether the causal relation can in some cases be perceived. An evaluation of theories of causality in chapter 3 suggested that the most viable account of causality was the conserved-quantity (CQ) theory. On the CQ theory, causal interactions are marked by the exchange of CQs between or among causal processes. If the CQ theory is correct, then causation is tied to real physical quantities, and one would expect that some aspects of causation can be perceived. What is required is that observers be sensitive to event dynamics, which involve energy, momentum, and other prime candidates for CQs. In chapter 4 we saw that the study of event perception has provided evidence that observers *are* in many cases sensitive to dynamics.

This is *not* to suggest that observers have privileged access to Newtonian ideas of space, time, mass, energy, or momentum. What it does say is that there are specific physical determinants to causal interactions, that these determinants are known quantities which have specific effects on the resulting kinematics, and that the data in real observational situations are rich enough to reconstruct at least the qualitative dynamics of the situation. If observers are picking up on the kinds of dynamic variables which are also CQs, then on the CQ theory, observers can be said to perceive causal interactions.

That much says nothing whatsoever about the phenomenological manifestation of such a perception. Specifically, again, it is not claimed that observers have an impression which corresponds simply to Newtonian mass or energy. The axes may be rotated, compressed, and otherwise distorted, so it will clearly not do to ask, "Did you see energy conserved?" just as it will not do to ask, "How *causal* did that appear to you?" *All* we can predict from the combination of CQ theory with event perception is that observers should be *sensitive* to exchanges of CQs. More precisely, they should be sensitive to *violations* of proper exchanges, which is to say, violations of relevant conservation laws, such as conservation of energy or momentum.

Therefore, otherwise inanimate displays which do *not* obey the conservation of energy ought to look unnatural, because the conservation violation *specifies a hidden cause.* Furthermore, even that should not look odd or notable if we believe in the basic tenets of event perception: that observers can *recognize* events according to their kinematics. Plenty of real events violate billiard-ball conservation laws. First, almost everything observers encounter is dissipative. Second, observers encounter animate motion all the time, without thinking it is unnatural. The reason is that these kinds of motions violate conservation laws in very *specific* ways, which allow them to be *recognized,* which is exactly what we would want out of a combination of a theory of causation and a theory of event perception.

Therefore, for these experiments, we chose a lawful kind of violation which in the context of the displays given to observers, nevertheless did not specify any kind of naturally occurring energy source or sink. Michotte had plenty of displays which specified impossible collisions, but because he made arbitrary variations, it is difficult to tell just what each display is specifying, and just what is the relationship between one display and another. As Runeson [94] demonstrated, some of Michotte's "noncausal" displays were perfectly natural collisions with different masses, or different properties. For that reason the experiments done here vary the displays according to some unnatural, but lawful, dynamic.

If observers are sensitive to conservation laws, and we have truly avoided the conflation of our events with other common natural events, then observer responses on naturalness rating tasks should indicate the extent to which they perceive causal relationships to account for the observed kinematics. Given the theories developed in chapters 3 and 4, observer naturalness ratings serve as measures of their implicit perception of causality. No claim is made that the naturalness ratings are the ideal instrument. Indeed, they are still subjective and prone to contamination by conscious deliberation. However, their established track record in the event perception literature indicates that they are superior to direct questions about abstract concepts like causation. If a theory can be developed for making use of reaction time data, or even better, if a perception-action task could be engineered to yield a measure of sensitivity to proper conservation and dissipation, that would be a step in the right direction.

Previous chapters have set out reasons for investigating a philosophical concept like causality with empirical methods. In short, causality is a physical concept which, whatever its physical manifestation and underlying reality, is fundamentally linked to the foibles of human perception. In fact, as noted in the opening epigraph, it *can* be debated whether causation is in the world or the equations, but there is no doubt that it is in our thinking. That link to our thinking has constrained the development of philosophical theories of causality—usually by appeal to simple untested intuition— and ultimately any adequate account of causality must account for our perceptions using the underlying relations and quantities in the world to which our perceptions are attuned.

In the more cognitive reasoning-and-judgment-oriented approach exemplified by Cheng's work [17, 16, 43], the interaction between philosophical theory and empirical testing has been very fruitful, and promises a powerful formalism for representing causal inference and discovery. However, the framework developed by Cheng relies on some prior, primitive notion of causation to fill in the background knowledge of possible causal relations. Those basic, mechanical kinds of causal interactions are loosely called "mechanisms," or "mechanism-based explanations." In related studies, Ahn *et al* [2, 1] have shown that humans prefer mechanism-style explanations to even more predictively-powerful covariational accounts. Human ideas of causality are tied to some notion of "mechanism," which the Ahn and Cheng accounts leave ultimately unexplained. I have attempted to extend their interactive approach between philosophy and psychology to explore the foundations of "mechanism," through the perception of physical causality in simple billiard-ball style situations where it is easiest to investigate the role of CQs in causal processes.

This work is a continuation of Michotte's attack on the problem of causal perception, but with more attention to the physical adequacy of the stimuli. It thereby draws on the event perception and naïve physics traditions as well. Of course, the original problem and approach belong to Hume. It was he who asked in what the causal relation consisted, and who sought the answer in psychological principles. Like Hume, I do not deny a reality to causal relations in the world, but I do think that the category of causation is most meaningfully appreciated as a perceptual or psychological construct.

However, I would not wish to identify one with the other. Although Hume failed to identify the relation, and ultimately we may do no better here, there are some events or combinations of events which evoke the idea of causation, and some which do not. There is no expectation of perfect accuracy. What the perceptual system would identify as a cause will likely miss some real causes and include some non-causes. Since on our understanding, the true causal interactions are those involving the exchange of real CQs, of which humans are almost certainly unable to perceive, human perceptions of causality will of necessity be approximations, in the way that Newtonian mechanics is a medium-scale approximation to more general theories.

What seems clear is that a great many causal interactions are perceivable, and that there are base cases about which intuitions are fairly certain. As Hume noted, observers have strong expectations of what will happen when one billiard ball collides with another. Indeed, they know much more than that it will cause the other to move. Although they can not articulate it very well, observers are fairly sensitive to *how* the ball should move, and can recognize incorrect motions. Billiards experts can predict trajectories with astonishing precision, even while maintaining *contradictory* conscious beliefs about the causes. [52] This point should be remembered: the fact that observers are aware of the underlying dynamics does not imply that they can articulate that knowledge or use it in any non-perceptual setting. And that is why a theory which holds the possibility of the perceptual sensitivity to physical quantities does not conflict with the long hard struggle that is the history of science. The fact that humans are in fact good enough at navigating the world to imply that we are sensitive to the actual physics and causal structure does not in the least imply that we *represent* that information in the correct way.

In summary, then, the motivating theory for the two experiments presented in this chapter is that the basic causal mechanisms which seem to underly other psychological accounts of causal attribution, explanation and discovery are something like Michotte's ampliation—understood not on his theory but as direct perceptual causality in the tradition of event perception—and this relation is founded in lawlike interactions governed by conservation and entropy, in a way that links up with the CQ account of objective causation.

That dynamic properties can be extracted from kinematics (and possibly other channels such as sound, deformation, vibration, *etc.*) allows us to test whether observers are sensitive to conservation principles. Rather than ask observers to give their impression of how "causal" something looked, I used naturalness ratings. If observers cannot notice that conservation violations are unnatural, then however important CQs may be to objective causality, they are not the properties determining causal perception.

I have run two experiments, designed to investigate sensitivity to violations of the conservation of energy and momentum. Experiment 1 investigated primarily the violation of the conservation of energy in a single-ball two-dimensional series of bounces. Experiment 2 presented two-ball one-dimensional collisions which violated the conservation of energy or momentum, or both. What has emerged out of both of these studies is that observers are rather strongly attuned to entropy. They are consequently partially attuned to anomalies in energy and momentum, but with an asymmetry that makes them relatively insensitive to violations that could be construed as frictional or inelastic losses. There does appear to be a perceptual arrow of time.

5.1 Overview

The bulk of this chapter consists of the formal presentations of each experiment, written according to the conventional style of published psychology papers. This brief section is intended to provide a quick take-home summary of each experiment, in order to supplement the detailed expositions which follow.

5.1.1 Experiment 1: Gravity bounces

Experiment 1 was conducted in three parts, not counting a small pilot study. In the first part, observers watched a ball bounce across a textured landscape, and had to rate the naturalness of the bounce on a scale from 1 (unnatural) to 5 (natural). The question under investigation was, if the conservation of energy is violated in an unusual way, will the observers notice? Will they be as sensitive to increases as they are to decreases? The "unusual way" was that gravity was made to vary with time over the course of the bounce. From the objective, simulation-based point of view, gravitational energy was either being pumped into the system (increasing gravity) or extracted from it (decreasing gravity).

Observers responded asymmetrically in an unexpected way. They were sensitive to decreasing gravity, but not increasing gravity. In the simulation frame of reference, that looked bad for CQ theory: there is no reason observers should be more sensitive to dissipation than to anomalous increases of energy. However, recall that the variation was chosen to be completely unnatural. Time-varying gravity specifies no kind of real interaction except itself, and observers have no experience perceiving fluctuating gravitational fields. From the *observer point of view*, the energy dynamics were much more complicated, and the results are in fact consistent with an asymmetry favoring sensitivity to energy increases, when those increases happen while the ball is rising, but not when it is falling.

Two followup tasks indicated that the asymmetry was not a matter real threshold-style sensitivity, but semantic indifference. When given a chance, observers demonstrated that they could *distinguish* both increasing and decreasing gravity from normal gravity, but that only decreasing gravity looked strange, at the levels used.

Experiment 1 seems to support the possibility for perception of violations of conservation laws in specific contexts, but not generally. It suggests that entropy is at least as relevant as conservation. This in turn also supports a more simple associationist account: observers attune to what they experience. However, the details are interesting since observers' perceptions ran counter to objective gains or losses in energy, but with phase-dependent *apparent* gains and losses.

5.1.2 Experiment 2: Collisions

The goal in experiment 2 was to test differential sensitivity to energy and momentum conservation using one-dimensional collisions. The technique was to have observers choose the most natural of a pair of collisions. One of the pair was a normal reference collision, and the other involved either a loss or gain of momentum or energy or some combination of both.

This experiment provided the most straightforward example of an event involving the exchange of conserved quantities, and observers were sensitive to changes, but varied greatly as to which changes were most noticeable. On average, a trend emerged. Observers were more reliably able to detect cases of increased energy or momentum, in keeping with the results of experiment 1. In fact some cases of lowered energy and momentum looked more natural than the reference case. Since momentum is never lost in a collision—even a completely inelastic one—that was surprising.

One explanation is that the reference case itself was highly elastic ($\epsilon = 0.94$), and the high elasticity made the reference case itself somewhat unnatural. Another explanation is that although momentum is never lost during a collision, momentum *is* lost throughout the event to friction. The experiment simulated air friction, but did not take sliding friction into account.

5.2 Experiment 1: Bounces

In Experiment 1 we followed the design used in Muchisky and Bingham [82] and McConnell, Muchisky, and Bingham [79] to present subjects with simulations of balls falling freely and then bouncing on a hard surface. While Muchisky and McConnell were interested in size perception from gravitational dynamics, we were interested in observer sensitivity to accumulation and dissipation of energy and momentum. Therefore, departing from their design, the ball was always simulated at the same distance, and was always the same size and mass. The bounce event occurred against a ground texture, with proper occlusion of background elements and by foreground elements. Experiment 1 was comprised of three variations. The first variation was the main experiment. Variations 2 and 3 were run as followup conditions in order to clarify an ambiguous result from variation 1.

5.2.1 Variation 1

The main goal was to judge observer sensitivity to a gradually-introduced conservation violation achieved by allowing gravity to vary over time. Were observers sensitive to this kind of manipulation at all? Were the equally sensitive to both additions and deletions of energy? As a baseline for comparison we also ran observers on simple conservation violations achieved by making the elasticity of the ball greater than 1, so that it exited from a bounce with a greater velocity than it entered. If anything were to look unnatural, that should. There were, therefore, two main sub-experiments in variation 1: elasticity and gravity. Either:

1. the ball's elasticity was varied, from 0.6 to 1.2, with gravity remaining earth-normal *or*
2. the gravitational field was made to increase or decrease during the course of the display, with elasticity always pegged to 0.9 (steel-on-steel).

Although the displays were intermixed, these manipulations were carried out separately. Each manipulation had five levels. In addition, to control for duration effects, all displays were constrained to last either 2.5, 3.5, or 4.5 seconds, resulting in a 3×5 design for the elasticity experiment, where the ball was always dropped from the same maximal height[1] of 3.5m. In the gravity experiment, elasticity was fixed at 0.9 based on a pilot study for what seemed the most natural.[2] For the case of varying gravity, we did not know whether, for a particular duration, it would be more important for subjects to witness a greater number of cycles, which required lower drop heights, or fewer but larger and more distinct cycles. Therefore there were two conditions:

- First, we kept the drop height constant and maximal at 4.5m and cut off the display whenever the time was up, as in the elasticity experiment. (fixed maximal initial height condition)
- Second, we varied the drop height so as to guarantee that each display had exactly five ground strikes in the allotted time.[3] (fixed number of bounces condition)

[1] Any higher and the bounce with elasticity = 1.2 went off-screen.

[2] This in itself is interesting. It is not clear why subjects should prefer an e of 0.9. Nothing short of a super ball or steel-on-steel has that kind of elasticity.

[3] We could not do this for the elasticity experiment because at low elasticities, only two or three bounces are distinguishable even for maximal drop heights. After that the ball just rolls.

As we will see, the fixed-height condition was much easier for subjects.

Method

Observers. Twelve graduate and undergraduate students at Indiana University participated in the study. All had normal or corrected-to-normal vision. The observers were paid $5 per hour.

Display generation. A heavy (6.8 kg) ball was simulated at a viewing distance of 50m. The ball appeared as a two-dimensional black outline with gray shading and a black radius line to indicate orientation and rotation. The background was a sagebrush texture gradient (see figure 5.1). QuickTime™ movies of simple planar events were created using event-dynamic models generated by the application Interactive Physics II[64], which has built-in facilities for movie generation and collision-monitoring. All models included components for gravity and air resistance. A Runge-Kutta algorithm was used to calculate positions at successive time intervals of $\Delta t_{calc} = 0.0015s$. Every tenth such position was used to generate an animation frame, so that $\Delta t_{display} = 0.015s$, or 66.7Hz. The application MacroMind Accelerator™ [77] was used to lock the frame rate of the simulations to the refresh rate of the monitor (66.7Hz), resulting in smooth animation without aliasing. Interactive Physics II does not simulate compression during bounce, so this was approximated by cutting off the bottom 2 pixels of the QuickTime™ movie, resulting in a flattening of the ball upon impact.

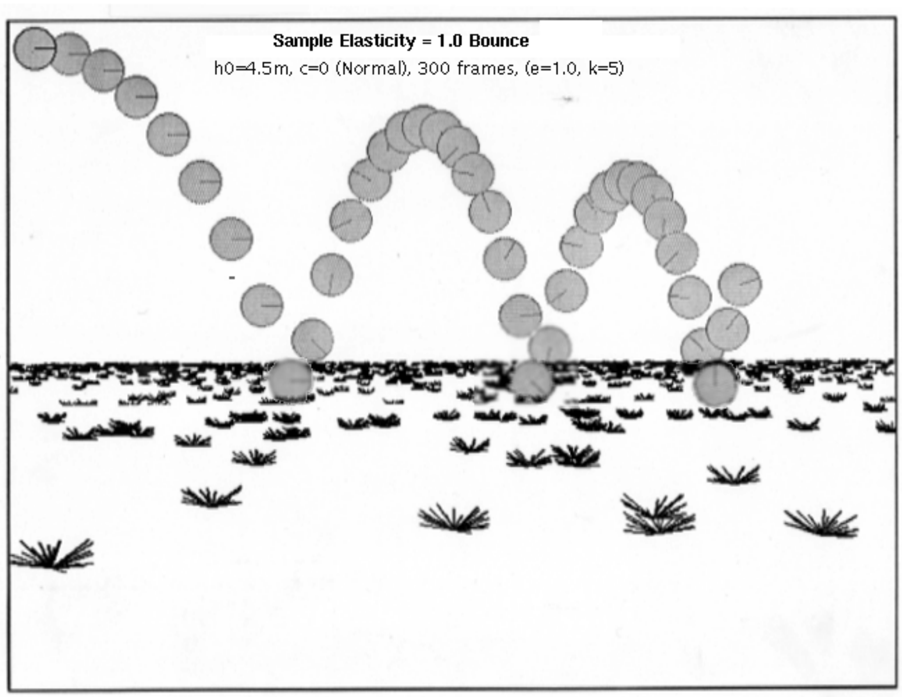


Figure 5.1: Display for Exp.1: An elapsed-time view of the bounce trajectory for normal gravity, perfect elasticity, and air friction proportional to velocity. Because of program limitations, the frames around the bounces were not generated and had to be pasted in by hand, and thus look fuzzy and do not show the proper occlusion.

Animations were generated in advance for each of the forty-five conditions investigated. In each display, the ball appeared stationary at its starting location, and began moving when subjects clicked the mouse. At the end of the movement, the ball disappeared. The sequence repeated three times in response to subject mouse-clicks. The whole set of 45 trials was presented in a predetermined random order, broken up into 2 blocks of 22 and 23 trials (each with 3 repetitions). Each subject saw three repetitions of the 45 trials, with each repetition using a different ordering of trials.

The displays used the full 24.8 × 18.5 cm (640×480 pixel) display of the Macintosh II, and subjects sat approximately 0.5m from the display.

Instructions. Instructions were presented one line at a time on screen in response to observer mouse-clicks, in a slide-show format. Observers were instructed to watch all three repetitions of each trial, and then to rate the naturalness from 1 (unnatural) to 5 (natural). They were told to pay attention to the form and timing of the motion. They were told that the rotation of the balls was always natural, and that the balls may get a little flat on impact but that was OK. The first four observers were also told to describe what seemed unnatural on those displays they marked unnatural and for which they had a clear impression why. This instruction may have put those observers in a more analytical mindset than was ideal. At any rate, it took a long time, so subjects five through twelve were told to perform that free-response task for only one of the sets of 45 trials. Observers were instructed that they could take breaks between each block (22 or 23 trials).

Procedure. Observers were seated in front of the computer with a mouse and an answer sheet and pencil in front of them. An architect's lamp illuminated the answer sheet and the overhead lights were shut off to reduce glare on the screen and help with the illusion that the subject was looking out through the monitor. Observers read the instructions and looked over the response sheets and had the opportunity to ask the experimenter any further questions. They then worked through a practice block to become familiar with the mechanics of the program and the range of naturalness of the displays. No feedback was given during the practice block. After answering any further questions, the experimenter then left the room. Subjects averaged about an hour. In a few cases the experimenter had to re-enter the room at subject request in order to restart the animation if it stopped.

Results

Elasticity study. We performed an ANOVA on the data from the thirteen subjects. There was a small but significant [$F(2, 20) = 3.54$, $p < .05$] main effect of Duration indicating that at shorter durations, subjects had a harder time discriminating natural and unnatural displays. More important, there was a significant effect of elasticity [$F(4, 40) = 18.76$, $p < .001$].

Figure 5.2 illustrates that subjects detected $e = 1.1$ and $e = 1.2$ as very unnatural. Although it appears to show a peak at $e = 0.9$, a Tukey HSD post-hoc analysis shows that the two points ($e = 0.6$, $e = 1.0$) are not significantly different from $e = 0.9$. Likewise, the points $e = 1.1$, and $e = 1.2$ are not significantly different from each other, although both are significantly different from all of the other points.

There was a significant two-way interaction between Duration and Elasticity [$F(8, 80) = 3.19$, $p < .005$]. Figure 5.3 shows that this interaction is chiefly due to the unnatural (high-elasticity) events looking even more unnatural at longer durations, as one would expect.

Gravity sub-study. An ANOVA with Order as a between-subjects vari-

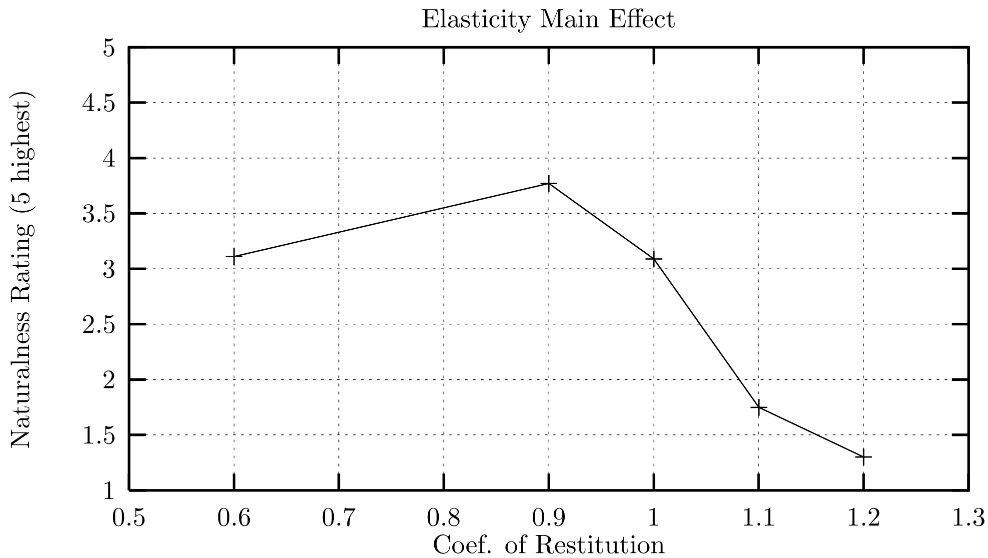


Figure 5.2: Ratings of Naturalness according to level of elasticity.

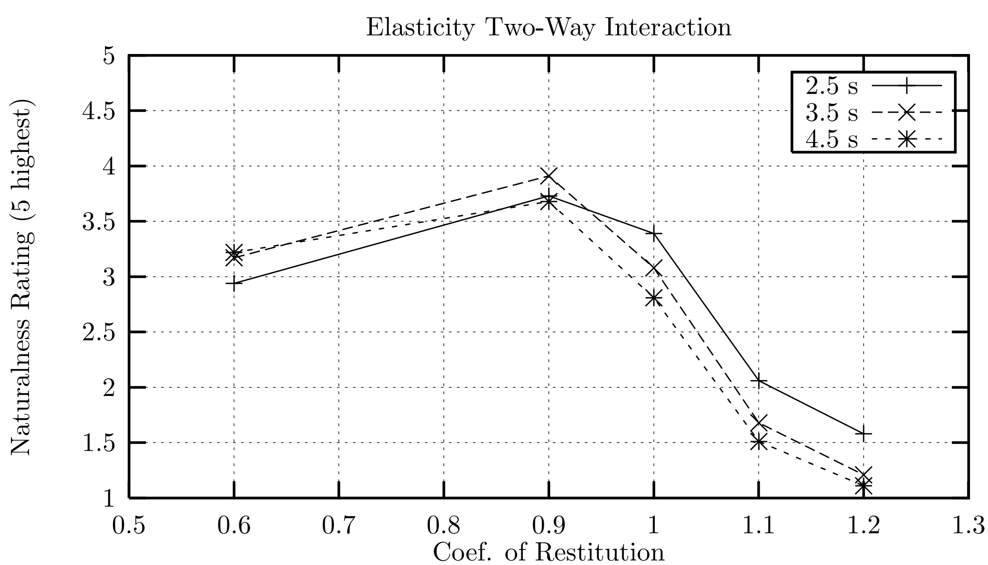


Figure 5.3: Two-way interaction between Duration and Elasticity

able turned up main effects of Gravity [$F(4, 40) = 29.71, p \ll .001$], Duration [$F(2, 20) = 14.67, p < .001$], and Fixed-condition (whether number of bounces or initial height was held fixed) [$F(1, 10) = 8.71, p < .05$]. There was a significant interaction between Gravity and Fixed-condition [$F(4, 40) = 22.74, p \ll .001$], and a significant two-way interaction between Gravity and Duration [$F(8, 80) = 3.65, p, .005$]. One three-way interaction between Gravity, Duration, and Order of presentation was evident [$F(8, 80) = 2.28, p < .05$].

The main effect of Gravity indicates that the displays simulating decreasing gravity appeared unnatural. Participants did not in these displays discriminate between normal gravity and increasing gravity. The main effect of Duration was expected. It indicates that shorter durations make discrimination more difficult. The main effect of Fixed condition indicates that subjects are better able to discriminate unnatural low-gravity motion when the initial screen height is fixed and large. The two-way interaction between Gravity and Fixed Condition shows very clearly how this plays out (see figure 5.4). In figure 5.4, the upper curve,

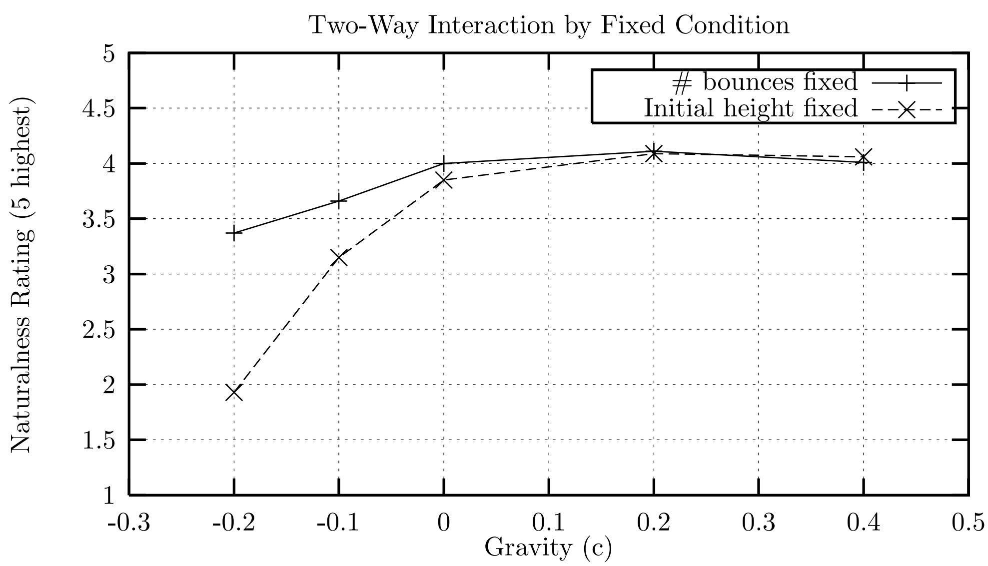


Figure 5.4: Two-way interaction between Gravity and Fixed Condition. Fixing the number of bounces rendered subjects less able to discriminate natural from unnatural motions, presumably because initial heights for $c < 0$ had to be quite small.

representing performance when initial height was not fixed, shows almost zero discriminability, as confirmed by a Tukey HSD post-hoc analysis: almost none of the points on the upper curve can be discriminated.

In contrast, on the curve depicting a fixed initial height, subjects did much better. It remains true that the right three points are indistinguishable. However, in the 4.5s condition, all three leftmost points can be discriminated from each other. Shorter-durations conditions have less discriminability.

In sum, the two-way interaction depicted in figure 5.4 demonstrates that fixing the number of bounces was an ineffective way to present these simulations. The problem was not the fixed number of bounces, but the corresponding reduction in initial height and resulting impoverished trajectories. Subsequent analyses and experiments will use only the fixed-initial-height condition.

The two-way interaction between Gravity and Duration falls naturally out of those two main effects. At short durations participants were insensitive to all gravity manipulations. At longer durations they became sensitive to decreasing gravity, but not increasing gravity.

One potential worry was that the number of bounces (which varied according to initial height and gravity) would be a better predictor than Gravity. A multiple regression of Naturalness Rating on Gravity, Duration, and Number of Bounces put this worry to rest. It yielded a highly significant effect of Gravity ($p < .001$), a significant effect of Duration ($p < .05$), and no significant effect of Number of Bounces ($p = .28$).

Discussion

Specifics. The performance difference between fixing the number of bounces and fixing the initial height (figure 5.4) demonstrated that trajectory form was more important than number of bounces, such that in the future we will fix the initial height to some large portion of the screen height.n

Is it possible however, that in the fixed-initial-height displays, the sudden disappearance of the ball was in itself inherently more unnatural, hence artificially enhancing the discriminability? Fortunately, the two-way interaction (see figure 5.4) precludes this possibility. Since the two-way interaction shows no difference between the two fixed conditions for normal to increasing gravity, the potentially more unnatural look of a mid-arc disappearance seems a poor explanation of the data. Apparently subjects followed instructions and ignored any oddness due to disappearance.

Regarding durations, recall that the gravitational acceleration varied smoothly with time, so that the longer the display, the more the gravitational field deviated from earth-normal. The maximum rate of decrease was scaled so that gravity approached but did not quite reach zero by the end of 4.5s. In this case we should expect that longer durations help subjects to discriminate natural from unnatural events, because the unnaturalness of the event itself increases with time.

Energy Conservation. The Elasticity manipulation results were consistent with sensitivity to increasing energy (negative entropy), and insensitivity to decreasing energy (normal entropy). Other studies [90, 9, 8] on event dynamics, time-reversed motions, and animate versus inanimate dynamics show similar biases against anti-entropic motion. Real collisions are inelastic, and therefore we expected participants more readily to notice unexplained sources of energy than unexplained sinks. As we just saw in the elasticity experiments, they did.

In the Gravity manipulations, however, the energy of the system was manipulated smoothly and continuously. From the point of view of the simulation, increasing the value of g increased the system energy as well as the velocities, but decreased the bounce-height, relative to normal gravity. Decreasing gravity had the opposite effect, although *successive* bounces were never higher (unlike $e \geq 1.1$).

For comparison, figure 5.5 shows the graph of energy versus time for three bounces: quickly decreasing gravity ($c = -0.2$), normal gravity ($c = 0$), and quickly increasing gravity ($c = +0.4$). Please note that this is the E versus t graph computed from the point of view of what was actually done in the simulation. Call this the simulation frame of reference. Within each series, there are discontinuities when the ball hits the ground, because $e < 1$. In normal gravity ($c = 0$), energy loss between bounces is due only to air friction, which is highest near the bounces when velocity is highest. In this case, total

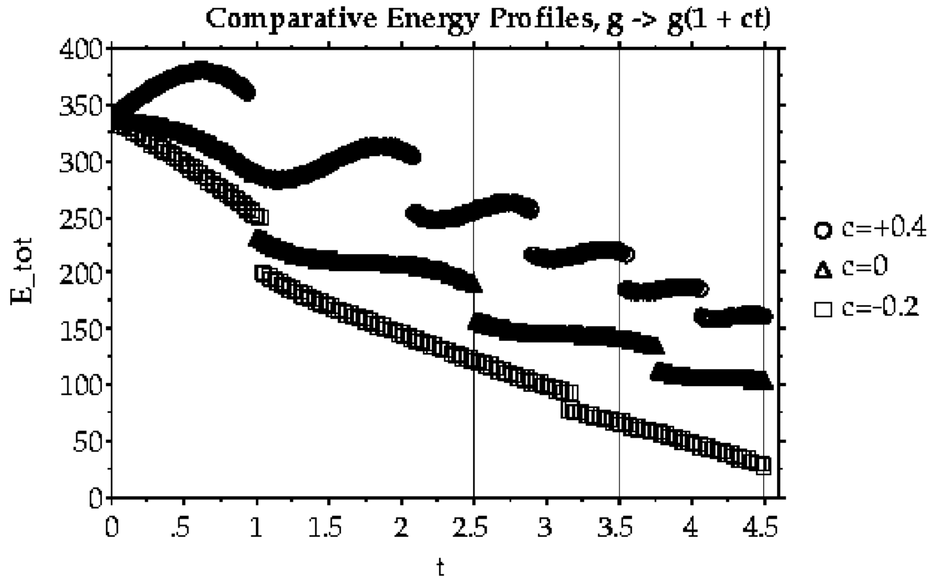


Figure 5.5: E_{tot} versus t for 3 principal values of c, according to the objective, or simulation frame of reference. Note that the hyper-gravity case ($c = +0.4$) is on top in this frame of reference.

energy decreases monotonically. When $c = +0.4$, energy *increases* between bounces, but not enough to offset losses at the bounce.

In the case of fastest increase of the gravitational field ($c = +0.4$), although the total energy over time still decreases, principally due to the inelastic bounce, the total energy between bounces increases for part of the cycle. Even so, we should note that losses due to air friction permit a net increase in energy only for the first two cycles. (Air friction losses are highest when the velocities are highest. On the curves, these are the regions with negative slope just after and just before each bounce, when the ball is moving fastest.) In the case of *decreasing* gravity, the potential field becomes weaker over time, resulting in an excessively rapid loss of energy. This may be seen in the bottom curve of figure 5.5.

So it appears that contrary to our expectation, participants were more sensitive to energy that was *decreasing* too fast. That result is contrary to the results from the Elasticity condition, and contrary to any expectations based on observation of conserved quantities and entropy. However, remember that figure 5.5 was computed in the simulation frame of reference. Participants, on the other hand, were not told that gravity could change. From their point of view, gravity was constant, and the energy of the bounce was changing in different ways. When we *recalculate* the energy values from Figure 5.5, this time forcing gravity to remain constant, the energy curves change drastically (Figures 5.6 and 5.7), in a way which makes more sense from the standpoint of a CQ theory of causal perception.

In the curve for $c = -0.2$ (Fig. 5.6), where gravity and simulation-frame energy decreased over time, there is instead a marked *increase* in apparent energy, but only during the *rising* phase of the cycle. This is because gravity progressively weakens, allowing the ball with its rebound velocity to achieve a greater height than it would have had gravity remained constant. The observer assuming a constant gravitational field would see this as anomalous. The apex

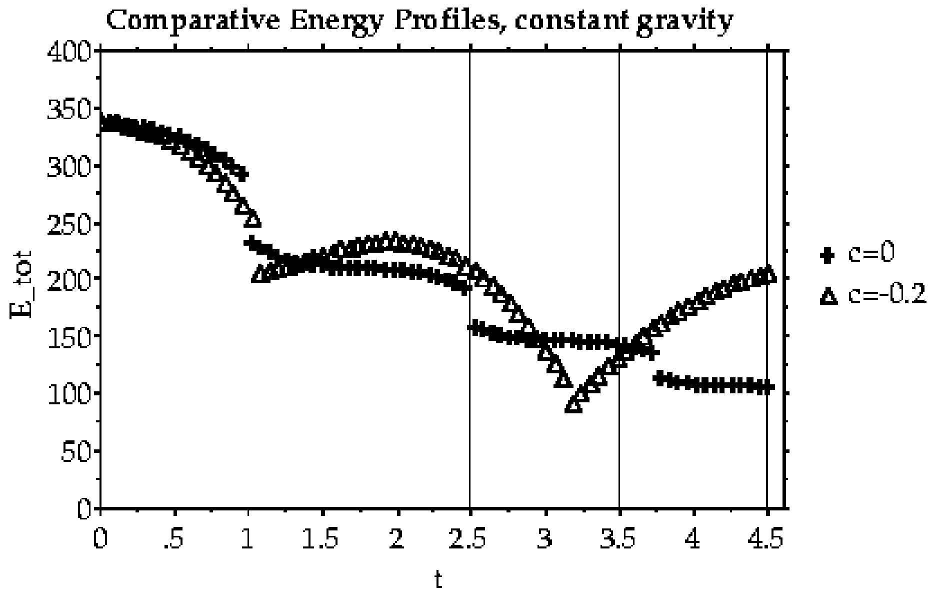


Figure 5.6: Recalculated graph of E_{tot} *vs* t comparing $c = -0.2$ and $c = 0$, assuming constant Earth-normal gravity. The curves for $c = 0$ have not changed and serve as a reference. The sampling frequency has been reduced in order to make the curves more legible.

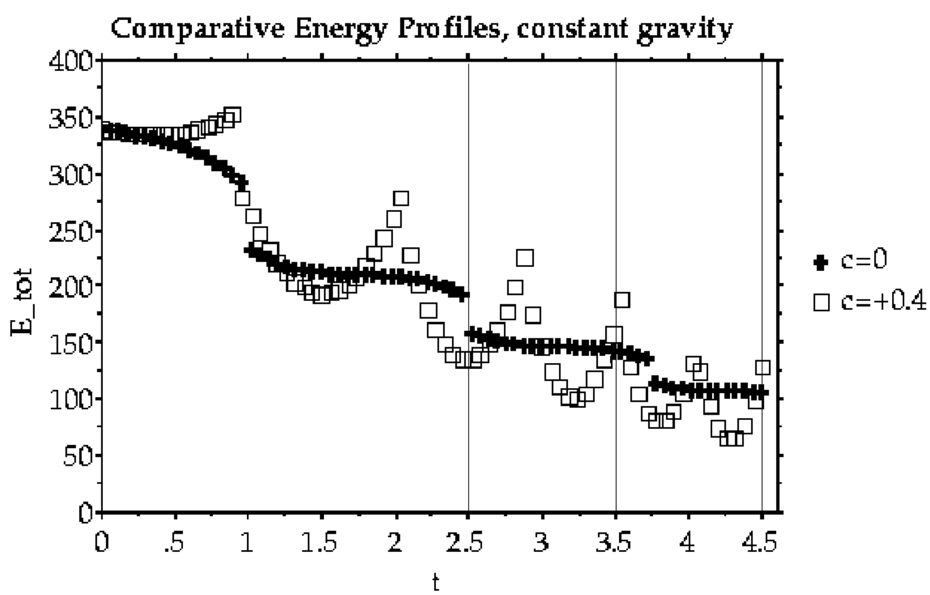


Figure 5.7: Recalculated graph of E_{tot} *vs* t comparing $c = +0.4$ and $c = 0$, assuming constant Earth-normal gravity. The curves for $c = 0$ have not changed and serve as a reference. The sampling frequency has been reduced in order to make the curves more legible.

heights never equaled the previous height (though some got close), but the *shape* of the energy curve is very different from the curves for Earth-normal conditions ($c = 0$). In short, there is a phase of increasing energy to which participants could attend.

The curve for $c = +0.4$ (increasing gravity) appears as a set of disconnected parabolas in Figure 5.7. Why were participants unable to attend to these significant increases and decreases in apparent energy?

Participants detected bounces whose energy gains occurred during phases where kinetic energy normally decreases: the rising part of the bounce. This is consistent across the elasticity and gravity conditions. Participants did not detect conditions where (kinetic) energy increased during descent or decreased during ascent. This was surprising, but entirely consistent with a strong role for a sensitivity to the direction of *entropy*. Detecting anomalous energy gains during descent is a much harder task because the signal is masked by the natural increase in velocity during descent.

Another possible explanation is the duration between bounces. Natural inter-bounce intervals decrease. Bounces under decreasing gravity have increasing intervals, marking them as unnatural, but bounces under increasing gravity merely have more rapidly-decreasing inter-bounce intervals. Since the simulated ball could have been of any material, any decreasing interval could have been possible. Indeed, if observers ignored the texture gradient and written information about the size and distance of the ball, and saw it as it was, about 2 cm in diameter and about a meter in front of them, then they would be right to judge smaller inter-bounce-intervals and shorter bounces as more natural. One sub-

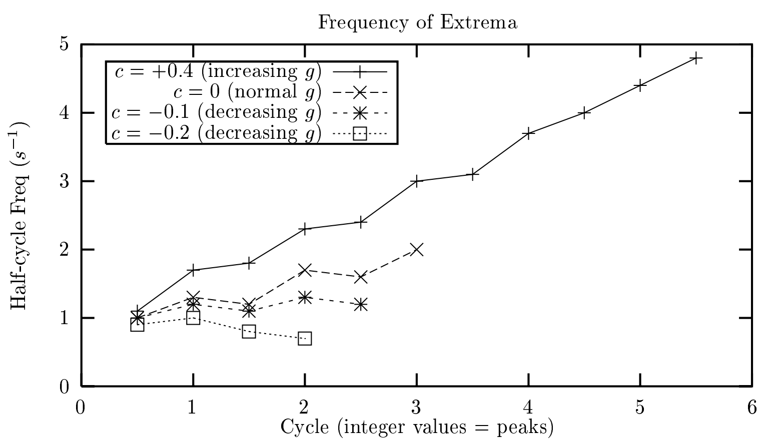


Figure 5.8: Graph of inter-bounce-interval (or frequency of extrema) for various values of c.

ject, a psychology graduate student and a musician, specifically mentioned that after awhile he began to judge timing between bounces, using decreasing inter-bounce intervals as a criterion for naturalness. This criterion of course made the bounces under increasing gravity look even better than the normal bounces. It is possible that subjects could have used just this timing parameter to make their judgments, although the spontaneous responses make it more likely that they paid attention to the motion trajectory. Free-response data suggests that most participants attended to trajectory form, but they could also have cued to inter-bounce intervals without realizing that some were too fast for the event they were told they were seeing.

In describing what seemed wrong about the decreasing gravity displays, subjects frequently said the balls were too "floaty," a description the experimenters agree was quite apt. This appears to describe the ball's tendency to go too high, and to be rather laconic in its subsequent descent, rather than just an indication of the inter-bounce timing. Even the subject who explicitly adopted a timing criterion said that in developing this criterion he had noticed that bounces he thought unnatural tended to have overly asymmetric trajectories.

Experiment 1 seems to support the possibility for perception of violations of conservation laws in specific contexts. When the total energy of the system increased during the rising phase of the bounce, subjects detected the anomaly. However, when energy increased on the descending phase, at least to the levels used in this experiment, subjects were unable to distinguish that motion from natural acceleration due to gravity. In no phase did subjects show sensitivity to excessive losses of energy.

So far we have seen that hyper-elasticity and decreasing gravity were discriminated from normal conditions, but increasing gravity was not. Was this asymmetry due to a real perceptual difference, or was it rather a matter of semantics? In other words, was it the case the increasing gravity was *indistinguishable* from normal gravity, or was it perfectly distinguishable but just not "unnatural"? If the latter, then the asymmetry is in part induced by the naturalness task itself. In order to answer the question and find out whether observers are unable to tell increasing-gravity cases from standard cases, we ran a followup experiment. The goal of variation 2 was to find out whether observers could discriminate between normal cases and increasing-gravity cases, even if they did not interpret the difference correctly.

5.2.2 Variation 2: Same-Different and 2AFC Task

In theory a fairly straightforward "same-different" task should have answered the question. Rather than rating naturalness, all observers had to do was tell whether two displays were of the same kind. Observers were instructed not to latch onto any particular peculiarities, but to judge on the basis of the overall motion. They were given feedback. After a session in that condition, subjects returned on a different day and were then given a 2-alternative forced-choice task without feedback. On the 2AFC task they had to decide which of 2 displays was the more natural.

Method

Observers. Five undergraduate students at Indiana University participated in the study. All had normal or corrected-to-normal vision. The observers were paid $7 per hour.

Display generation. As before, a 6.28 kg, 0.54 m ball was simulated at a viewing distance of 50 m. Elasticity was always 0.9, and only the fixed-duration 4.5s (300 frame) displays were used, since those produced the best performance in variation 1. Displays were either $c = +0.4$, $c = +0.2$, $c = 0$, $c = -0.1$, or $c = -0.2$. The initial heights of the displays were randomized, coming from the Low range (3.3 - 3.8m) or the High range (4.3 - 4.8m). Each pair had one "High" and one "Low," in random order. Likewise, the initial horizontal velocity was randomized: Low (1.0 - 2.0 m/s), High (3.0 - 4.0 m/s). The direction of motion for each pair (from the left edge moving right or conversely) was also randomized. All displays were generated and ordered in advance.

Instructions. Instructions were presented slide-show style as before. On day 1 (discrimination task), observers were instructed to check "Same" or "Different" after watching a pair of bounces, and to indicate how sure they were. On day 2 (the 2AFC task) they were told to choose which one looked more normal. As before they were told to pay attention to the form and timing of the motion and not to fixate on minor details. In addition they were told that the initial positions and velocities were slightly randomized to prevent them from being able to cue off of specific conditions. On day 1 they were told that they would receive feedback. On day 2 they were told that they would not.

Procedure. Observers were seated in front of the computer and made their answers on printed answer forms, as before. Alternate observers were counterbalanced for order of presentation. 12 blocks were presented in an hour or so: 6 blocks of decreasing gravity, 6 of increasing gravity. Each block had 12 trials (pairs), which is to say 4 complete runs through the 3 possible pairs. So after a day, each cell had $4 \times 6 = 24$ data points.

Because the task is now very clearly a discrimination or distinguishability task, a measure of distinguishability is appropriate for our analysis. The data were analyzed using d' measures from signal detection. The theory behind using signal detection measures for psychology experiments is set forth in Macmillan and Creelman's 1991 book, *Detection Theory: a user's guide* [76]. In addition, excellent software is available for simple or complex d' analysis [46]. Basically, d' is a measure of the discriminability of two signals. Graphically, it is the distance between the mean of one stimulus and the other stimulus used in the experiment. Each stimulus is assumed to be normally distributed about its mean, whether due to actual variation in the parameters, or assumed noise in the perceptual system. A $d' = 0$ indicates that the signal and the noise distributions lie right on top of each other, and are not reliably distinguishable. The larger d', the farther apart the means, and hence the more distinguishable the two signals. For ball-park assessments, it is convenient to take $d' \geq 1.0$ as a cutoff value for meaningful discriminability. However, in psychology it is the relative d's which are more relevant. Two d' values can be compared and checked to see if they are significantly different using the framework of standard statistical hypothesis testing.

Results

The day 1 d' data are presented in table 5.1 and figure 5.9. Recall that in variation 1 observers rated hypo-gravity as unnatural, but not hyper-gravity. The purpose of variation 2 was to find out whether observers nevertheless could *distinguish* hyper-gravity cases from normal cases, when the task did not involve explicit naturalness judgments. If we confine our attention to the second half of table 5.1, which is the performance for increasing-gravity displays, observers appear to be doing reasonably well. That portion of the table indicates that observers can distinguish the displays. However, the first half of table 5.1 indicates that observers were unable reliably to distinguish hypo-gravity displays, which is exactly the reverse of the result from variation 1. In fact, none of the observers achieved more than threshold (75% correct) in the hypo-gravity condition, and all observers were at chance for recognizing same-same displays as "same" on day 1 in the hypo-gravity condition, and only slightly above chance for the hyper-gravity condition.

The day 2 data show a different pattern (see table 5.2 and figure 5.10). There is a clear difference between discriminability for hypo-gravity and hyper-gravity, and observers can do the hypo-gravity task reasonably well, but cannot do the hyper-gravity properly.

	H	F	$p(c)$	($\pm$)	d'	$\pm$
Hypo-Gravity (collapsed over $-.1$ & $-.2$)						
Ma	.67	.58	.59	(.66)	.79	(.62)
Ad	.71	.58	.61	(.64)	.97	(.63)
Ro	.54	.50	.53	(.68)	.51	(.61)
Ra	.73	.54	.64	(.63)	1.19	(.63)
An	.77	.50	.68	(.59)	1.47	(.64)
Hyper-gravity (collapsed over $+.2$ & $+.4$)						
Ma	.81	.67	.65	(.62)	1.01	(.66)
Ad	.85	.42	.76	(.50)	1.99	(.66)
Ro	.67	.33	.67	(.60)	1.62	(.64)
Ra	.60	.46	.58	(.66)	.98	(.62)
An	.85	.33	.79	(.45)	2.23	(.68)

Table 5.1: d' and % correct ($p(c)$) data for Variation 2, day 1 (Same-Different). $n = 24$ (Same-Same), $n = 48$ (Same-Dif). Error bars ($\pm$) are 95% confidence intervals.

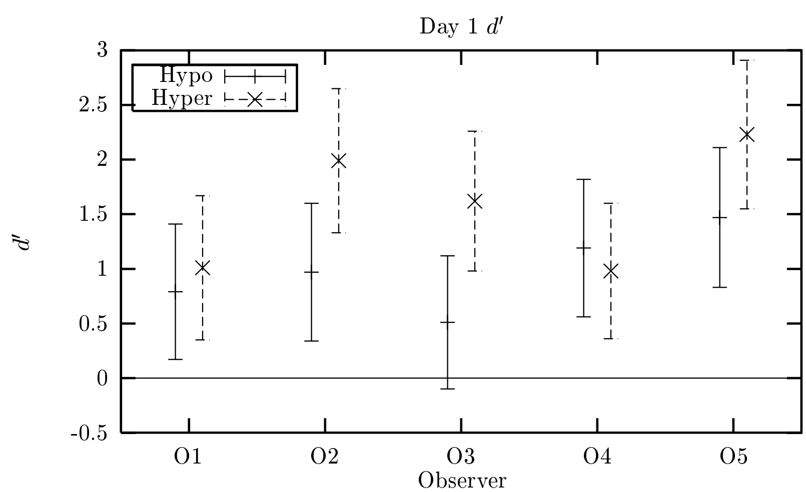


Figure 5.9: d' for Variation 2 Day 1.

Discussion

The interpretation that emerges from the variation 2 data is that observers did not understand part of the "Same-Different" task. Although that should have been the easier task, in fact it was quite ambiguous, because "same" did not mean "identical"—for no displays were identical—but "same kind." As noted, on day 1 performance was unexpectedly low on a task we already know from variation 1 that observers can perform. Much of the poor performance is explained by the fact that observers remained at chance on judging "same-same" trials: they did not understand what was being asked and had to guess.

Consequently, hypo-gravity performance was better on the no-feedback task on day 2, which means the separate tasks did not quite perform their intended

	H	F	$p(c)$	$(\pm)$	d'	$\pm$
Hypo-Gravity (collapsed over $-.1$ & $-.2$)						
Ma	.71	.29	.71	(.84)	.78	(.74)
Ad	.94	.11	.92	(.21)	1.99	(.86)
Ro	.75	.01	.83	(.38)	1.46	(.74)
Ra	.83	.06	.89	(.27)	1.81	(.83)
An	.81	.31	.75	(.51)	.97	(.64)
Hyper-gravity (collapsed over $+.2$ & $+.4$)						
Ma	.38	.75	.31	(.89)	-.70	(.75)
Ad	.47	.36	.56	(.67)	.20	(.59)
Ro	.42	.69	.36	(.63)	-.51	(.60)
Ra	.03	.94	.04	(.11)	-2.48	(1.07)
An	.44	.61	.42	(.66)	-.30	(.58)

Table 5.2: d' and % correct data for Variation 2, day 2 ("Which is more natural?" 2AFC) $n = 36$. Error bars ($\pm$) are 95% confidence intervals.

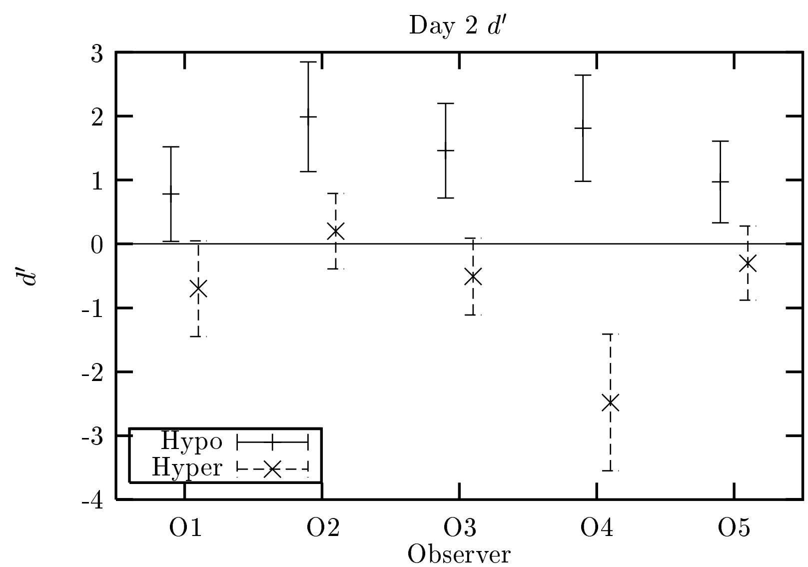


Figure 5.10: d' for Variation 2 Day 2.

functions. Nevertheless, the data do show some of the expected pattern. If we look only at the hyper-gravity data, we find that on day 1, given only the comparative same-different task, observers were able to distinguish hyper-gravity trials from normal trials with d's ranging from 1.0 to 2.2. However, after that very solid discriminability performance, when the question was switched on day 2 to a comparative naturalness rating, the discriminability falls apart. Four of the observers had d's not significantly different from 0, while one observer evidenced nearly perfect discriminability, with the semantics *completely reversed*: observer "Ra" clearly identified the hyper-gravity trials each time they occurred, and thought they looked more natural than the normal trials. At the same time, all observers were now able to distinguish the hypo-gravity trials from the normal trials, although performance levels varied.

However, because of the unreliable results for the hypo-gravity task on day 1, we ran another variation designed to address the shortcomings of variation 2. In variation 3, the instructions and task were made yet more straightforward. The purpose was to verify the basic design of variation 2.

5.2.3 Variation 3: Simplified Same-Different and 2AFC Task

Method

Observers. Two observers were run, a graduate student and a non-student in his mid-twenties. Observers were paid \$6 per hour, with a \$20 incentive reward for best performance, which went to aforesaid graduate student. All observers had normal or corrected-to-normal vision.[4]

Display generation. The displays for variation 3 were a subset of those from variation 2. Intermediate values of c were removed, so observers were tested only on $c = +0.4$, $c = 0$,and $c = -0.2$. Instructions were presented as before, but an "example" slide show was included. This showed several cases that were to be called "same" and contrasted them with extreme cases of "different," so that it would be clear to observers the kind of overall differences we were interested in.

Procedure. As in variation 2. Order of presentation for both observers was the same on both days. Six blocks of decreasing gravity were followed by 6 blocks of increasing gravity. Each block within a gravity condition presented the same 24 trial pairs, in different orders. Letting "N" stand for "Natural" and "U" for "Unnatural," each block had 6 <NU> and <UN> pairs, and 12 <NN> pairs. Therefore, after a day, each <NN> cell had $12 \times 6 = 72$ data points, as did the combined <UN> and <NU>.

Results

The results were very clear. On the day 1 same-different task, both observers had very high d's for both decreasing and increasing gravity conditions. On

[4]In fact, another graduate student was also run, but the experimenter inadvertently answered a design question in the observer briefing on day 1. The observer revealed that with the revealed information (that all displays were the same duration), she had deduced that any changes in the overall motion and timing would correlate with the number of bounces. Consequently, she reached ceiling almost immediately, and therefore did not benefit any from the feedback, nor did she exhibit any difference between day 1 and day 2. Indeed, in free response she explained the explicit response category mapping she used on day 2 so that she could still treat it as the day 1 task. In discussion she said the design was very similar to those she used in her own research on rhythm and timing perception. She won the \$20 hands down.

the day 2 "Which is natural?" task, the observers had high positive d's for the hypo-gravity cases, and high negative d's for the hyper-gravity cases (see figure 5.11. Therefore, the proper way to interpret the asymmetry in variation 1 is that observers were or would have been able to *distinguish* hyper-gravity displays from natural displays, but that nevertheless these displays did not appear unnatural. Observers were sensitive to the difference but unaware of its semantic content. This result confirmed the result obtained in variation 2, without the unfortunate glitch in observer response to the hypo-gravity condition.

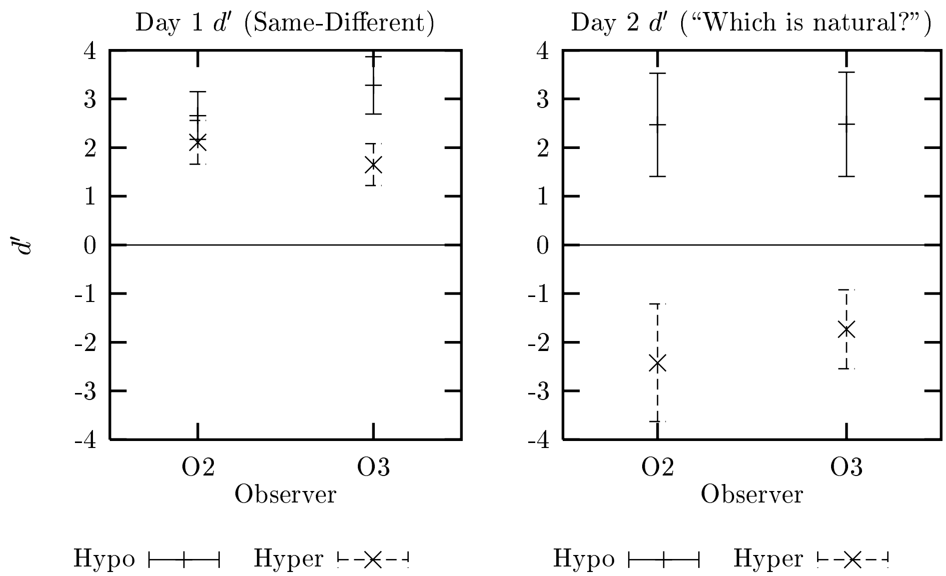


Figure 5.11: Exp. 1d d' data

5.2.4 General Discussion

In the main part of experiment 1, observers showed an asymmetric sensitivity to energy violations (see figure 5.4), prompting the question, "why is it asymmetric?" First we had to establish whether the apparent insensitivity to increasing gravity was a true indistinguishability or merely a question of what looked natural or unnatural. In variations 2 and 3 we determined that observers could discriminate between hyper-gravity bounces and normal gravity bounces, at the levels used in variation 1, but that the hyper-gravity bounces just did not look unnatural. In fact, experiment 1 established that the hyper-gravity bounces looked *more* natural than the reference case.

In order to explain the results in terms of the CQ theory, we transformed the graphs of energy versus time to the observer frame of reference, where gravity was not changing. Having done this it is possible to interpret observer behavior in terms of conservation and dissipation. When apparent energy gains happened while the ball was rising, these were seen as unnatural. The phenomenology is that the ball is decelerating too slowly, and going too high. Apparent energy gains on the descent did not look unnatural. Likewise, excessive losses of apparent energy on the ascending part of a bounce did not look unnatural. The result is consistent with a sensitivity to the direction of dissipation, or entropy, which also manifests itself in bounce frequency.

It is easy to explain the sensitivity to energy gains on the ascending arc: such gains oppose the loss of velocity and height which are normally part of dissipation, and the discrepancy becomes most apparent when the velocity is low, making small changes more salient. Conversely, on the descending arc, the velocity is already increasing. The fact that it is increasing somewhat faster is relatively more difficult to tell, especially since the effect was largest when the velocity is largest, which is to say when the resolution of the perceptual system is poorest. In addition, despite the extra energy, the ball still gets only as far down as the ground, so there is no height cue. Finally, excessive losses of energy on the rising arc are once again consonant in direction with the usual dissipation, differing only in magnitude, making any fine discrimination more difficult.

The asymmetry noted in variation 1 showed itself to be tied to the semantics of naturalness and unnaturalness, as we would expect from the CQ theory: although in straight same-different comparison tasks observers will be able to tell the difference in all cases, when placed in the context of naturalness, certain kinds of changes will be much more salient. The fact that the hyper-gravity bounces in some cases looked more natural than the reference case can be explained by noting the following:

- From the observer point of view, hyper-gravity bounces specified significant, if phase-dependent, energy losses
- Energy losses are expected in natural motion
- The reference case lost as little energy as was consistent with a natural bounce, making the bounce height and timing profiles of the hyper-gravity cases more representative of real bounces.
- The reference case used a large, heavy ball very far away, with dynamics appropriate for that case. But it looked like a very small ball right there on the computer screen. The hyper-gravity bounces were more consistent with the real dynamic of a small ball within an arm's length of the observer than the reference case dynamics were.

Therefore, experiment 1 supported the possibility for perception of violations of conservation laws in specific contexts, but not generally. It also suggested that entropy is at least as relevant as conservation itself. As discussed earlier, observers are used to dissipative systems, to the extent that perfectly conserved motion looks unnatural. The direction of causation is tied, in perception, to noticeable dissipation, suggesting that anti-entropic motion should look unnatural in part because it specifies backwards causation. Even when observers could detect the difference between hyper-gravity and normal, the apparently more dissipative hyper-gravity displays were preferred by naïve observers. Experiment 2 examined the differential sensitivity to energy and momentum conservation.

5.3 Experiment Two: Collisions

In experiment 2 we examined the effect of violating conservation laws in one-dimensional collisions between two bodies of equal mass but unequal initial velocities. Kaiser and Proffitt [62] studied observer precision on various kinds of momentum violations in two-dimensional collisions where the second ball was initially at rest.

The goal in experiment 2 was to test differential sensitivity to energy and momentum conservation using one-dimensional collisions. The technique was to have observers choose the most natural of a pair of collisions. One of the pair was a normal reference collision, and the other involved either a loss or gain of momentum or energy or some combination of both.

5.3.1 Method

Observers. Eleven adults participated in the study, including ten graduate and undergraduate students and one staff member at Indiana University. All had normal or corrected-to-normal vision. The observers were paid $6 per hour for two 1-hour sessions. A $20 reward was offered for the observer with the best performance.

Parameter Generation. Completely elastic collisions obey both the conservation of momentum and the conservation of energy. Thus, if the masses are m_1 and m_2, the incoming velocities are v_1 and v_2, and the outgoing velocities are u_1 and u_2, the collision equations are:

$$m_1 v_1 + m_2 v_2 = m_1 u_1 + m_2 u_2 \tag{5.1}$$

$$m_1 v_1^2 + m_2 v_2^2 = m_1 u_1^2 + m_2 u_2^2 \tag{5.2}$$

Real collisions are not perfectly elastic, and do not obey the conservation of (kinetic) energy. However, even perfectly inelastic collisions conserve momentum. If we define the elasticity ϵ as the ratio of separation velocities of the two balls (positive = receding from each other, negative = approaching), real collisions are governed by the conservation of momentum and the definition of elasticity, ϵ:

$$m_1 v_1 + m_2 v_2 = m_1 u_1 + m_2 u_2 \tag{5.3}$$

$$\epsilon = \frac{u_2 - u_1}{v_1 - v_2} \tag{5.4}$$

This is the governing set of equations used (albeit in two dimensions) in Interactive Physics [64, 65], and in previous work on the perception of collision dynamics [94].

In this experiment we will violate, independently, the conservation of energy and momentum, and not always dissipatively. If all we wanted to do was to violate the conservation of energy, the task would be trivial. Just as one can set $\epsilon < 1$, one can set $\epsilon > 1$, which indeed we did for the elasticity condition in experiment 1, variation 1. However, it is more difficult to violate only the conservation of momentum, in either direction. First, the available solution set begins to shrink as one must introduce imaginary numbers to solve the equation pair. Second, the justification of using ϵ depends on momentum being conserved. Equations 5.3 and 5.4 are insoluble for scale changes in momentum.

Instead, experiment 2 employed the following modified equation set in place of the real collision laws.

$$\gamma(m_1 v_1 + m_2 v_2) = m_1 u_1 + m_2 u_2 \tag{5.5}$$

$$\Lambda(m_1 v_1^2 + m_2 v_2^2) = m_1 u_1^2 + m_2 u_2^2 \tag{5.6}$$

When γ and λ are both 1, this is just the conservation of momentum and energy, equivalent to the real collision equations with $\epsilon = 1$. As shown in figure 5.12, the correspondence in predicted exit velocities is still quite good for the reference case used in experiment 2, in particular for the mass ratio used, which is 1:1. Of

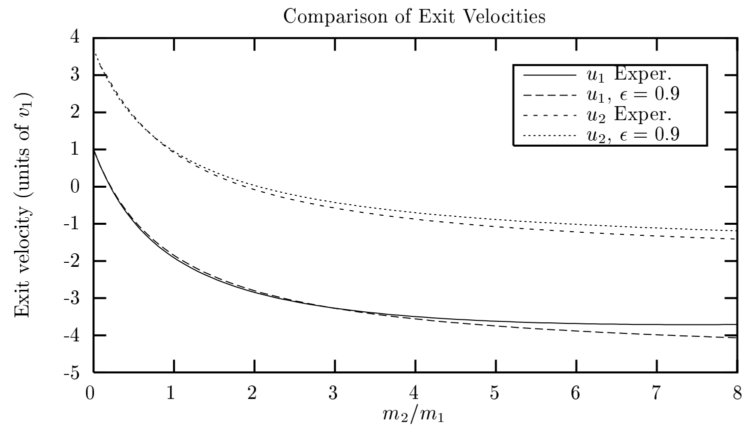


Figure 5.12: Plot comparing u_1 and u_2 from the canonical collision equations at $\epsilon = 0.9$ and the modified equations when $\Lambda = 0.9$, for various mass ratios $\frac{m_2}{m_1}$, and for $\frac{v_2}{v_1} = -2$.

course, the point of the new equation set is that it is now possible to scale the post-collision energy by the factor λ and the post-collision momentum by the factor γ. There are still limits. As γ deviates from 1, the range of incoming velocities for which a non imaginary solution can be found drastically narrows. In fact we can take the fact that the collision equations are only generally solvable (with nonimaginary numbers) when momentum is conserved as an argument in favor of the fundamental nature of CQs. Nevertheless, when $\gamma = 2$, the range of nonimaginary solutions is wide enough to include a collision between relative incoming velocities of $+1$ and -2.

The closed form of the solution becomes complicated to work with, but we can ask Mathematica to solve equations 5.5 and 5.6 for us. We find that $u_1 =$

$$\frac{m_1(m_1v_1 + m_2v_2)\gamma \mp \sqrt{m_1m_2(-(m_1v_1 + m_2v_2)^2\gamma^2 + (m_1 + m_2)(m_1v_1^2 + m_2v_2^2)\Lambda)}}{m_1(m_1 + m_2)} \quad (5.7)$$

and that $u_2 =$

$$\frac{m_2(m_1v_1 + m_2v_2)\gamma \pm \sqrt{m_1m_2(-(m_1v_1 + m_2v_2)^2\gamma^2 + (m_1 + m_2)(m_1v_1^2 + m_2v_2^2)\Lambda)}}{m_1(m_1 + m_2)} \quad (5.8)$$

Since by convention m_1 is the object entering from the left, we are only interested in the "-" part of the $\mp$ in the equation for u_1, and the "+" part of the $\pm$ in the equation for u_2, since these are the terms that correspond to a rebound from a collision.

Once again, the real and substitute equation sets correspond perfectly for the case of a perfectly elastic collision, and quite closely when $\gamma = 1$ (conservation of momentum) and $\Lambda = 0.9$ (slightly inelastic), as shown in figure 5.12. However, the ratio of kinetic energy, Λ is not the same as the elasticity ϵ, and the difference is more apparent as the ϵ departs from 1. For instance, in a perfectly inelastic collision ($\epsilon = 0$) for two objects of identical mass, $v_1 = 1$

and $v_2 = 0$, conservation of momentum yields an exit velocity for the combined object of 0.5. The energy ratio, Λ is then 0.25, not zero. For collisions used in this experiment, where the relative speeds are $v_1 = 1$ and $v_2 = -2$, the energy ratio for a completely inelastic collision ($\epsilon = 0$) would be $\Lambda = 0.1$. Of course, the curves for u_1 and u_2 will depart radically from the canonical curves when we allow γ to be something other than 1, violating the conservation of momentum. That is, of course, the point. The deviations when momentum is not conserved are sufficient to guarantee that there are no real collisions, even of objects with different masses, which would fit the given velocity profiles. However, observer sensitivity has a limited resolution, and it is not known if some of the unnatural collisions might be close enough perceptually to some real collision to effectively specify a possible mass ratio.

The equations were solved for two identical masses with relative initial velocities of $v_1 = 1$ and $v_2 = -2$ for the combinations of Λ and γ shown in table 5.3. For comparison to parameters from the canonical equations, table 5.4 shows the

		Λ 0.45	0.9	1.8	
	1/2	-1.28	-1.73	-2.36	u_1
		0.78	1.23	1.86	u_2
γ	1	-1.44	**-1.91**	-2.56	u_1
		0.44	**0.91**	1.21	u_2
	2	-1.35	-2.12	-2.87	u_1
		-0.65	0.12	0.87	u_2

Table 5.3: Resulting velocities u_1 and u_2 (relative to v_1) for the 9 combinations of Λ and γ used in the experiment. The reference case of a natural collision in marked in bold text.

effective elasticities, ϵ expressed by the velocity ratios. For convenience we will

		Λ 0.45	0.9	1.8
	1/2	0.69	0.99	1.40
γ	1	0.62	**0.94**	1.37
	2	0.24	0.75	1.25

Table 5.4: Resulting effective elasticities ϵ for the 9 combinations of Λ and γ used in the experiment. Note that for these velocities, the elasticity of the reference case is 0.94, slightly higher than 0.9.

later refer to the cells in this table by box number according to the following grid,

1	2	3
4	**5**	6
7	8	9

where energy (Λ) increases across the top and momentum (γ) increases down the side.

Initial v		At collision				Exit collision	
v_1	v_2	v_1	v_2	γ	Λ	u_1	u_2
12	-24	10.9	-21.8	0.5	0.45	-14.0	8.5
					0.9	-18.8	13.4
					1.8	-25.7	20.2
				1	0.45	-15.6	4.7
					0.9	-20.9	10.0
					1.8	-27.9	17.0
				2	0.45	-14.8	-7.0
					0.9	-23.1	1.3
					1.8	-31.3	9.5
9	-18	8.0	-16.0	0.5	0.45	-10.2	6.2
					0.9	-13.8	9.8
					1.8	-18.8	14.8
				1	0.45	-11.5	3.5
					0.9	-15.3	7.3
					1.8	-20.5	12.5
				2	0.45	-10.8	-5.2
					0.9	-16.9	0.9
					1.8	-23.0	7.0

Table 5.5: Actual screen velocities (cm/s) for the two principal velocity conditions used. Actual velocities are generated from the single table of relative velocities given in table 5.3. The velocities at collision are slightly less than initial velocities because of air friction. Exit velocities are computed from the velocities at collision.

Using the *relative* exit velocities given in table 5.3, *actual* velocities were chosen for the simulation. There were two overall speed levels and each speed level could have the faster ball come in from the left or the right. The simulated balls were 2 cm in diameter. In all, this made for $9 \times 4 = 36$ displays. In the 2AFC task, displays from the boxes on the edge of the grid were paired with box 5 from their same grid, making $8 \times 4 = 32$ pairs, $\times 2$ for order so 64 total pairs in a full run through the stimuli.

The array of actual screen velocities (in cm/s) for the two overall speed levels is given in table 5.5. Since air friction was simulated, the velocities at collision were slightly less than the chosen initial velocities. Exit velocities were calculated from the actual velocities at impact. Sliding friction was not used, but would have had more of an effect on the velocities at collision.

Display Generation. The balls were two black textureless disks 2cm in diameter shown on a plain white background. At the start of the display, both balls were offscreen. Within a few frames, the slower ball entered from its side at its initial velocity. When it was halfway across, the faster ball entered from the other side, at twice the speed. The balls collided in the middle of the screen and rebounded. The display ended when the first ball to leave the screen had done so.

QuickTime™ movies were generated from Interactive Physics 5[65], but without using the built-in collision engine. Instead, balls moved according to an initial velocity with losses due to air friction up until the frame of the collision, and then with the exit velocities from table 5.5, minus losses due to air friction. Frictional losses were slight. Collision frame was determined visually, and starting positions were tweaked slightly so that in the collision frame the two circles just abutted.

The simulation generated frames for every $\Delta t_{display} = 0.015s$, so they would run frame-for-frame on the monitor's refresh rate of 66.7Hz. The application MacroMind Accelerator™ [77] was used to ensure such rapid playback and to lock the frame rate of the simulations to the refresh rate of the monitor (66.7Hz), resulting in smooth animation without aliasing. Accelerated animations were generated in advance for each of the thirty-six initial velocity conditions. The four reference cases (boxes 5, normal collisions with $\gamma = 1$ and $\Lambda = 0.9$) where then paired with each of the other eight cells in their own matrix, in both orders, so that the total number of pairs in the stimulus set was $8 \times 4 \times 2 = 64$. The experiment was then assembled in MacroMind Director(TM) version 3.1 (the last version to support Accelerator) under control of a Lingo script, so that display order was shuffled on the fly within each block, and observers responded on-screen immediately after each pair was shown. Displays were broken into three blocks: the 16 pairs with Λ at 0.9 (boxes 2 and 8 paired with 5), the 16 pairs with γ fixed at 1 (boxes 4 and 6 paired with 5), and the 32 remaining pairs where the non-reference member of the pair had both $\Lambda \neq 0.9$ and $\gamma \neq 1$ (boxes 1,3,7, and 9 paired with 5).

The displays used the full 24.8 cm (640 pixel) width of the Macintosh II display, and subjects sat approximately 0.5m from the display.

Instructions and Procedure. Instructions were presented on screen as the experiment began. Observers were instructed to watch the pair of collisions, and decide which one they thought was more natural, and to rate how sure they were. They were also told on day 1 that this first task would take about 30 minutes after which they should take a short break, and then proceed to the second task which would be a straight rating of the naturalness of a single collision. On the second day observers were told that the whole hour would be devoted to judging between pairs. The experimenter stayed in the room until they had seen one or two of the four practices, and answered any questions about the mechanics of responding, and encouraged them to take short breaks between the blocks if needed. Overhead lighting was left on. No feedback was given during the practice or the experiment.

In the 2AFC task (first half of day 1, all of day 2), observers watched a pair of collisions, and then had to choose which one was more natural, and rate their confidence. This was accomplished by having them press 1 (definitely first), 2 (probably first), 3 (possibly first), 8 (possibly second), 9 (probably second), or 0 (definitely second). On the straight judgment task (second half of day 1), they had to rate the naturalness from 1 (very unnatural) to 6 (very natural).

At the end of the trials on day 2, observers were given a short questionnaire which asked how difficult they found the task, what cues or rules they adopted, which displays were easier and why, what they thought was going on, and how many years of undergraduate psychology and physics they had.

5.3.2 Results

2AFC task

As in the case of variations 2 and 3 of experiment 1, the task at hand is essentially a discrimination task, so d' measures are appropriate. We computed eight d' measurements for each observer, one for each pairing of the other 8 boxes with box 5 (collapsing over the four different initial velocity configurations). In this case the d' is measuring the relative ability of the observer to detect the unnatural collision in the randomly-ordered presentations. The results are listed in table 5.6 and shown in the scatterplot in figure 5.13.

	Box Number							
Observer	1	2	3	4	6	7	8	9
O1	-1.24	0.05	2.77	-0.84	2.77	0.49	-0.68	2.32
O2	0.25	-0.61	-1.03	1.21	-1.06	1.84	1.95	0.41
O3	-0.19	-0.41	-0.89	0.85	0.18	1.02	1.18	0.49
O4	-0.64	-0.27	1.19	-0.30	1.73	0.92	1.20	1.09
O5	0.56	-0.49	0.04	2.37	0.23	1.59	2.63	0.61
O6	-0.20	-0.10	0.81	0.33	0.83	0.28	0.50	0.56
O7	-0.50	0.33	1.50	-0.31	1.48	0.49	0.49	0.89
O8	-0.96	0.10	2.74	-0.62	1.90	0.91	1.24	1.99
O9	-0.19	0.00	1.74	-0.68	1.64	1.02	0.87	1.82
O10	-1.58	0.30	2.32	-1.39	2.07	0.27	0.24	2.43
O11	-0.61	-0.32	1.49	1.02	1.97	1.02	2.95	2.32
Pooled	-0.44	-0.17	0.87	0.09	1.02	0.83	0.89	1.16

Table 5.6: d' values by box and subject. The 95% confidence intervals on each d' are roughly ± 0.59 around $d' = 0.9$, ± 0.55 around $d' = 0$ and ± 1.18 around $d' = 2.6$.

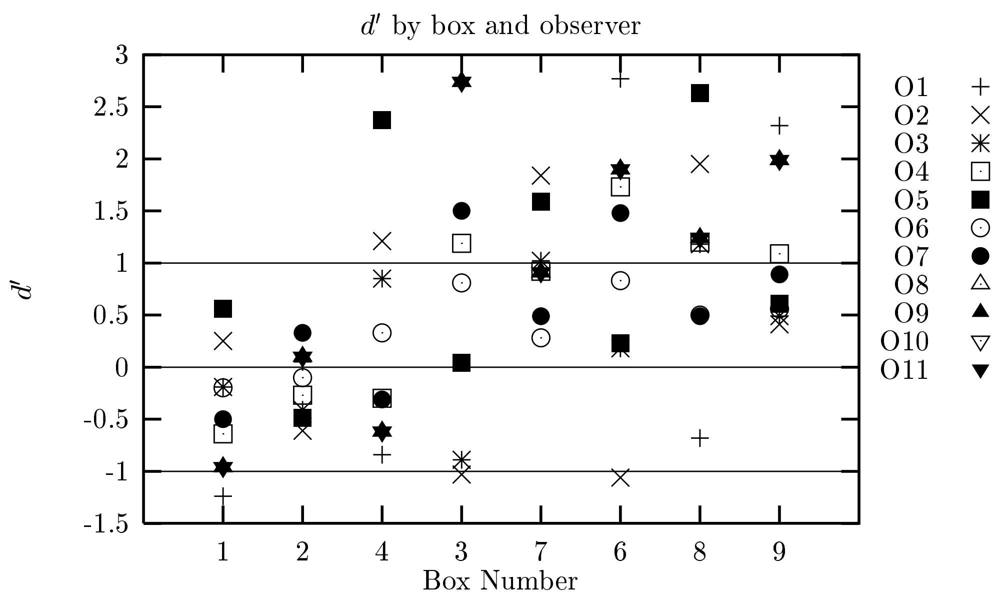


Figure 5.13: The assorted d' values obtained for each observer for each box. Arranged roughly by increasing Λ and γ.

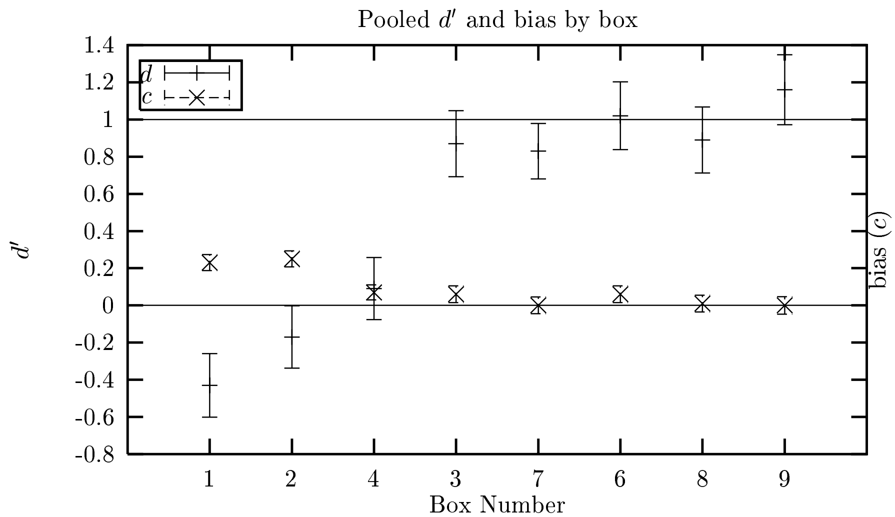


Figure 5.14: *Pooled* sensitivity and bias for all observers, according to box. Arranged roughly by increasing Λ and γ.

The pooled d', listed in table 5.6 is plotted separately in figure 5.14 along with the bias measure c. The plot shows standard 95% confidence intervals for a pooled d' computation. However, those confidence intervals only apply to the mythical observer with the pooled data. In fact, as figure 5.13 shows, inter-observer variance is much higher. The following pairs of points from the pooled data are not significantly different from each other: (1,2), (3,7), (3,6), (3,8), (7,6), (7,8), (6,8), (6,9). All other non-self pairs are. Most notably, the cases of pure energy increase (3,6) cannot be distinguished from cases of pure momentum increase (7,8), although their combination (9) can be distinguished from almost everything.

Both plots demonstrate that observer sensitivity increases roughly with increasing Λ and γ, proceeding from box 1 diagonally down to box 9. Furthermore, the pooled data shows a distinct difference between the top left third (boxes 1, 2, and 4) and the lower right two-thirds. Furthermore, when performance is high, overall bias is very close to zero, but when performance is low, overall bias is higher. The measure of bias, $c = -0.5(z(H)+z(F))$, where z is the z-transform which takes a probability to the corresponding z-score of a normal distribution (0 for Pr = 0.5, 1.96 for Pr = .95), is a measure of observers' tendency to say "Natural."

However, because observers used a six-level rating task, we have the ability to test for variations in d' at different confidence levels. The six-category task implicitly asks observers to alter their biases for each category, so we can use the rating data to plot ROC curves, and estimate d' from the ROC curve. Since the rating task had six categories, there are five data points for each curve. The ROC curves require more data to be meaningful, and so we have only computed them for the pooled data. ROC curves were fitted using Rscore+ [46], which uses procedures from [76]. They are shown, by box, in figures 5.15 and 5.16.

Random guessing generates equal hit and false response rates for all levels, and thus defines the chance line, which is the main diagonal running from (0,0) to (1,1) on the ROC graphs. Lines above and to the left of the chance line indicate increasing sensitivity. Lines below and to the right indicate increasing sensitivity in the wrong direction. Curves which are symmetrical about the

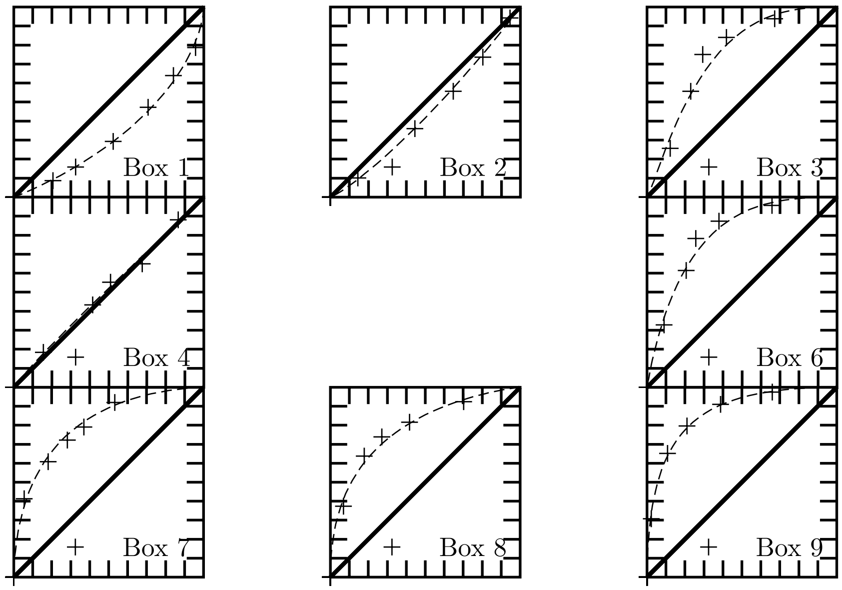


Figure 5.15: ROC curves by box. Pooled data, 11 observers. Same data as figure 5.16.

minor diagonal are measured adequately by d'. Others, such as the curve for box 3, are noticeably asymmetrical; the sensitivity for these boxes needs to be estimated from the ROC curve.

Box	Original d'	ROC-based d'
1	-0.448	-0.43
2	-0.142	-0.17
4	+0.038	+0.09
3	+0.661	+0.87
7	+0.885	+0.83
6	+0.849	+1.02
8	+0.800	+0.89
9	+1.131	+1.16

Table 5.7: Comparison of d's from using the rating data (ROC curve) with the original estimates.

The d' measures derived from the ROC curves are compared with the standard d's in table 5.7. Most of the estimates are quite close, but those for boxes 3 and 6 are noticeably different. We would expect these to be the most different since they have the most asymmetric of the ROC curves. Once again, ROC curves that are symmetric about the minor diagonal will generate the canonical d'. The revised d's are plotted figure 5.17. All data points are significantly different from each other except the following pairs: (1,2), (7,6), (7,8), (6,8).

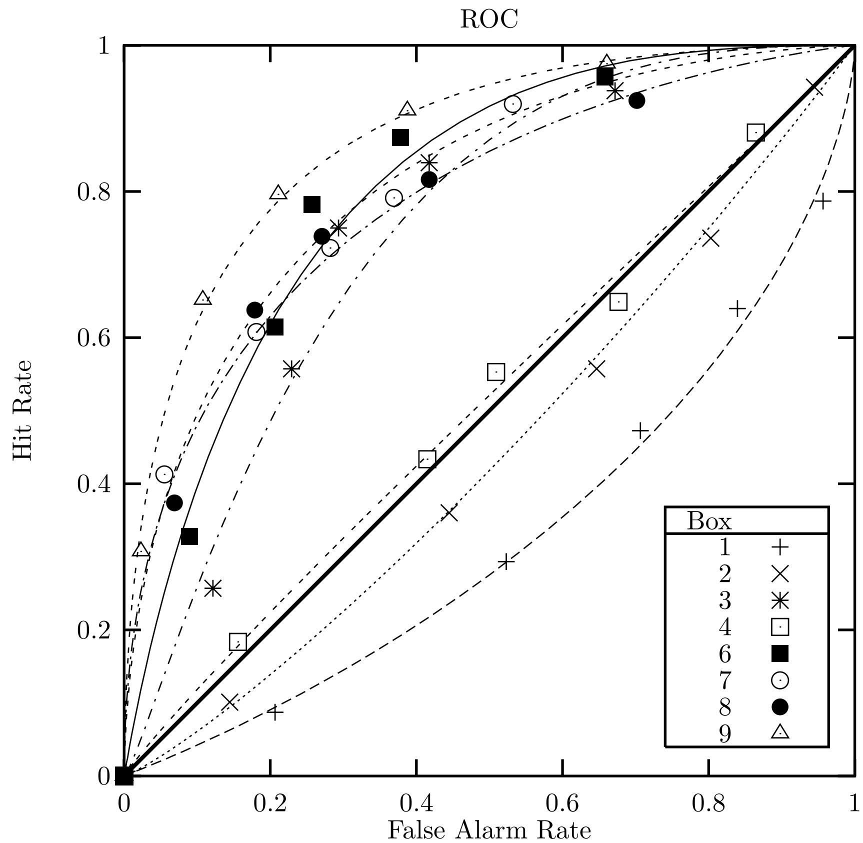


Figure 5.16: ROC curves presented together. Pooled data, 11 observers, same as figure 5.15. Horizontal axis is false-alarm rate (0..1). Vertical axis is hit rate (0..1). Tic marks are every 0.1.

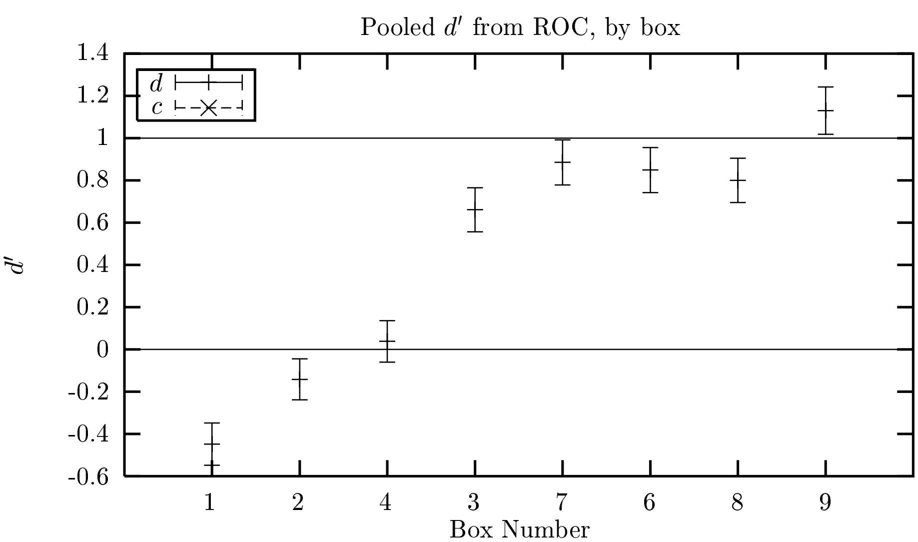


Figure 5.17: d's calculated from the ROC curve

Once again, there is evidence of a clear break between the upper left corner of the matrix and the lower right two-thirds.

Single judgment task

As a check on the task, in addition to the forced-choice response, on day 1 each observer judged the naturalness of single displays on a six-point rating scale, without feedback. Each observer made 24 judgments on each box, 1 through 9.[5] We then computed d' measures for every other box compared with box 5. First, in parallel with the standard d's computed in figure 5.14, the d's generated from collapsing the rating data into a binary "yes/no" task are presented in figure 5.18.[6] What is most interesting about this graph is that for the first time we are seeing a difference in sensitivity between excess momentum and excess energy. In the single task, the high-momentum conditions (7 and 8) appeared more unnatural than the high-energy conditions (3 and 6), and the combined condition now was not distinguishable from the other high-d' conditions. Second, an ROC analysis was run on the 6-category rating task. The model produced the d' values presented in figure 5.19. Finally, for comparison, the four pooled d' plots are reproduced in one graph in figure 5.20.

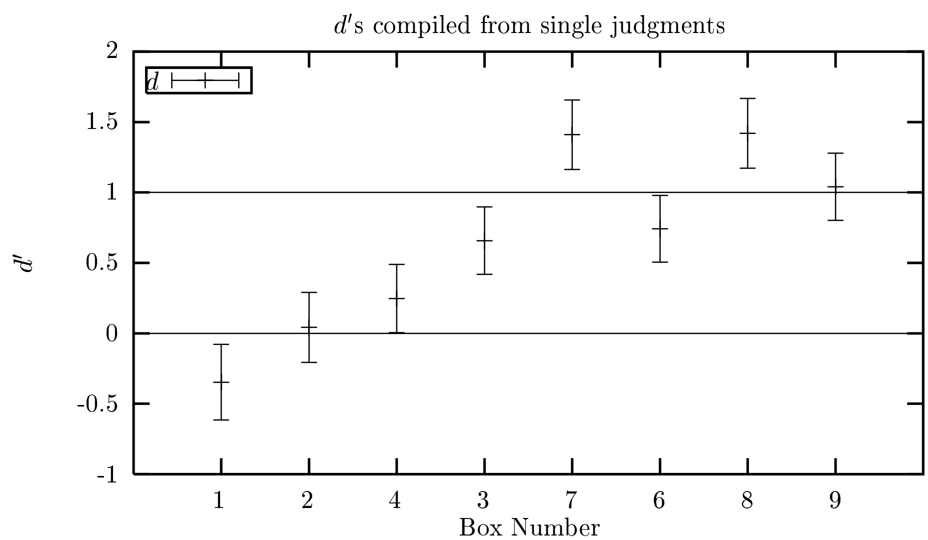


Figure 5.18: d' for single judgment "yes/no" task.

As is most clear in the combined figure (figure 5.20), the pooled data all follow the same basic pattern of increasingly accurate discrimination with increasing additions of energy and momentum. In addition, all analyses indicated that box 1 was perceived as *more* natural than the reference collision. The single-display rating task corroborates the results of the forced-choice task, despite large differences between observers. The most notable difference between the tasks occurred in box 7, which was judged much more unnatural in the single-display rating task. Box 7 is the one case where both balls left the screen

[5] The data file from observer 1 was lost, so the resulting dataset is composed of the observations of 10 observers, for 240 observations per box..

[6] The biases are not shown in figure 5.18 because every box is being compared to box 5, so the hit rates are artificially fixed, and therefore the biases ($c = -0.5(z(H) + z(F))$) would just track the false-alarm rate. In figure 5.19, there is a different c for each box, since the demands of the task require observers to change their biases.

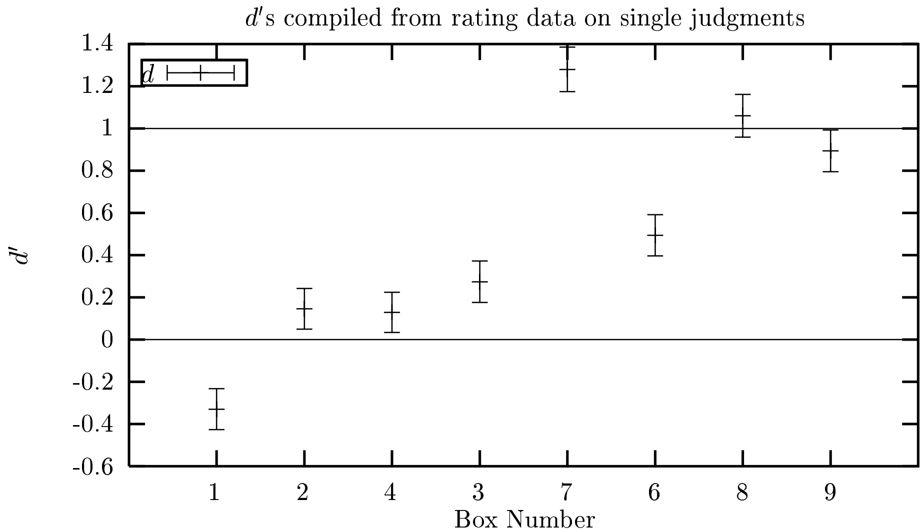


Figure 5.19: d for single judgment six-category *rating* task.

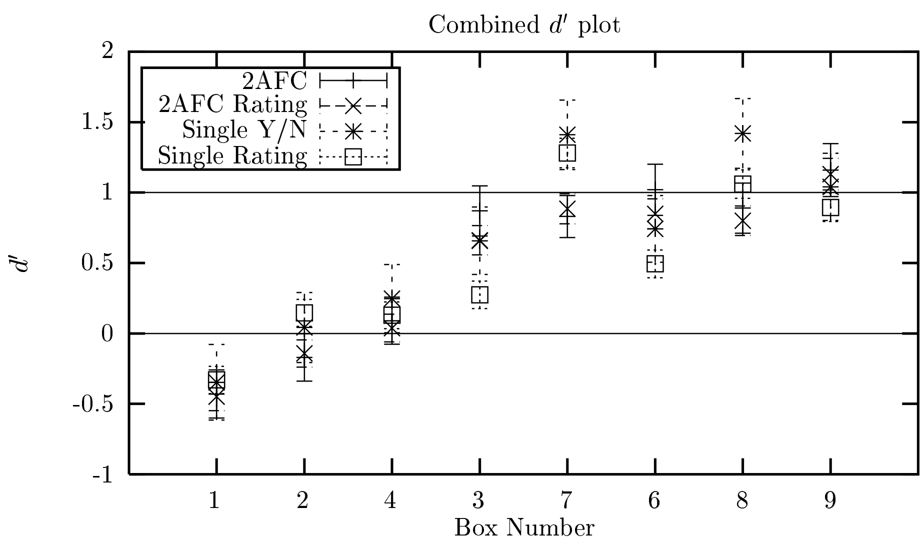


Figure 5.20: Combined d' from the four pooled data conditions.

in the same direction, so it is not surprising that it stood out, although it is interesting that it did so only while rating displays singly, and not when comparing naturalness.

In fact, in the single-display rating task, the high-momentum conditions (7 and 8) appeared more unnatural than the high-energy conditions (3 and 6), possibly including the combined condition. Figure 5.20 reveals that the difference is almost entirely due to the higher d's for boxes 7 and 8 in the single-display task.

5.3.3 Discussion

The principal result was that combinations of high energy and high momentum were detected as unnatural. Conversely, when both energy and momentum were lost, the displays were reliably considered to look *more* natural than the reference case. Box 4 was on average indistinguishable from the reference case, as it should have been. Boxes 4 and 5 both portrayed normal collisions with conserved momentum ($\gamma = 1$) but different elasticities corresponding to the different values of Λ (0.9 and 0.45). Those boxes with normal momentum but high energy were not distinguished from those cases of normal energy but high momentum. The box representing increase of both energy and momentum (9) was distinguished from everything else.

The results are entirely in accord with experiment 1. Observers detect that increases in conserved quantities are unnatural, but actually prefer decreases. In the case of decreased momentum this is somewhat strange, since momentum is always conserved in inert collisions. However, in the events *surrounding* a collision, momentum is dissipated to the environment, and it may be that observers were sensitive to the overall small loss of momentum. It is hypothesized that in an experiment which takes surface friction into account, the significant negative d's would disappear.

Although the pooled results conform nicely to experiment 1, individual observers varied widely. Some were clearly sensitive to energy increases but not momentum increases. Others were sensitive to momentum increases but not energy increases. A few even achieved significant d' scores, in both directions, on box 4. Much of the variance is probably caused by the partly cognitive nature of the task. Several observer reported noticing a particular kind of event, but having explicitly to decide what to call it. At the point of such explicit reasoning, observers are no longer just reacting to the percept. Unfortunately, this experiment could not escape ratiocination. For instance, it *is* possible for two balls to move off in the same direction after a collision, if one is more massive than the other. It is hoped that pooling the data averages out any arbitrary choices explicitly made by individual observers.

Although the overall trend was stable across different tasks and analyses, in the single-display rating task, cases of excess momentum were rated more unnatural than cases of excess energy. One explanation is that when given the opportunity to run through the whole dataset and explicitly rank the unnaturalness of displays in comparison to each other, observers did in fact find that violations of momentum conservation were more salient. Phenomologically, this is certainly possible since in box 7, both balls exited in the same direction, and in box 8, the faster ball came almost to a stop after the collision. These kinds of displays were qualitatively different from the others, and afforded easy category boundaries for recognition and rating. By comparison, in the principal task, observers were only able to compare two displays, an unnatural and a natural one, and these displays were presented in blocks: momentum violations with energy fixed at the reference value, energy violations with momentum fixed at the reference value, and the corner boxes 1,3,7, and 9, where both energy and momentum were not at the reference value.

One deficiency of this experiment is that observers were never asked to choose which of *two* unnatural displays looked more natural. We can expect that such a 2AFC task would corroborate the results of the single-display rating task in showing momentum increases to be more unnatural than energy increases, but the combinatorial explosion involved in testing all of the cross possibilities prevented us from running that task in this experiment. Indeed, it would almost have to be run separately since there would be a full 56 cells even without duplicating any of the pairings studied in this experiment.

Once again, however, we found an asymmetry. This experiment provided the most straightforward example in this thesis of an event involving the exchange of conserved quantities. Observers were sensitive to changes, but varied greatly as to which changes were most noticeable. On average, a trend emerged. Observers were more reliably able to detect cases of increased energy or momentum, in keeping with the results of experiment 1. Also in keeping with experiment 1, some cases of lowered energy and momentum looked more natural than the reference case. Since momentum is never lost in a collision—even a completely inelastic one—that was surprising.

One explanation, as in experiment 1, is that the reference case was highly elastic ($\epsilon = 0.94$), and the high elasticity made the reference case itself somewhat unnatural. Another explanation is that although momentum is never lost during a collision, momentum *is* lost throughout the event due to friction. The experiment simulated air friction, but did not take sliding friction into account. If this explanation is correct, then by rerunning experiment 2 with realistic sliding friction, the significant negative d's would disappear. There would be two reasons for this. First, the reference case would not appear slightly unnatural in itself. Second, the loss of momentum would not mimic natural dissipation, since the natural dissipation would already be present in all cases.

Chapter 6

Conclusion

Hume's destructive arguments left behind a philosophical void. He had tried to excise the portions of Newtonian science which could lead to objectionable metaphysical conclusions, particularly about God. However, those arguments also barred the possibility of causal inferences, which in fact are relevant for science. Hume acknowledged that causal inferences occur all the time and are necessary not only for science but survival. However, his positive account was not up to the task of replacing what had been lost. Consequently, Hume left to both philosophy and psychology the problem of formulating a replacement account of causality.

On the philosophy side, the major contenders for most of the century have been regularity accounts, as can be seen clearly by their representation in the anthologies by Brand [10] in 1976 and Sosa and Tooley [104] in 1993. However, in the Sosa and Tooley volume, the process theory has moved to much greater prominence than it commanded in 1976. Since 1993, with the collaborative formulation of the conserved quantity (CQ) theory by Dowe and Salmon, the process theory has emerged as the most promising candidate in the field. Given its apparent compatibility with propensity theories and causal modeling (see for example, [59]), some version of the CQ theory is very likely the correct account of causality.

One particularly promising feature of the CQ theory, and indeed likely the reason it has gained acceptance, it that it seems to connect with everyday experience. Part of my argument has been to show how it is that CQ theory addresses both objective and perceived causality, and by linking both together makes sense of each. On the face of it, there is no reason that observers should be able to perceive the CQs of the universe. In fact, however, because CQs are not just properties which happen to be conserved, but fundamental dynamical properties which govern the interactions of objects, these are prime candidates for the kinds of properties a well-adapted organism ought to be attuned to.

Indeed, psychologists working in event perception have shown that observers are sensitive to dynamic quantities such as mass and momentum. It is important to realize that no claim is made that CQs are perceived *as such*—indeed, otherwise science would have progressed much faster than it has—but only that observers are sensitive to the underlying dynamics behind the perceived motions. This is possible because a rich enough set of trajectories, sounds, *etc* is specific enough to admit of only a few or perhaps only one possible underlying dynamic, given the constraints of the perceiver's environment.

If some causal interactions are perceivable, then we have an answer to Hume's *psychological* challenge as well as his ontological challenge. The fact that the union of empirical and philosophical approaches can yield a single coherent answer to a longstanding problem ought to recommend the approach. Philosophy does not always advance by paying attention to science. Sometimes it has to be wise enough to know that the relevant science cannot at that stage provide a reliable answer. However, I have argued that now is an appropriate time for philosophers interested in causation to incorporate ideas from event perception. In particular I offered a slightly amended set of definitions for the CQ theory which I think will facilitate integration with perception and hopefully, with causal modeling approaches.

As stated in the introduction, the purpose of this thesis was to defend the following three claims:

- that causation does not involve necessary connection
- that causal connections can sometimes be perceived, and
- that a proper study of causation requires attention to the empirical details of perception.

In the first two chapters I set out the case against the necessity view. In addition to the standard counterexamples, I tried to undermine Hume's motivation for continuing to use the necessity account, and argue that it was inconsistent to have done so given his definitive refutation of the whole idea of necessary connection in the physical world. Finally, I developed the CQ theory and showed how it avoids the issue entirely, while maintaining the essential and desirable features of an account of causation. Given the viable alternative, there is no need to remain with a necessity account, particularly in the face of mounting evidence of the pervasiveness of indeterminism in the world.

Towards the second claim, I showed why one would expect organisms to be able to perceive causality given the CQ theory of causation. Then I reviewed evidence from event perception that in fact observers are sensitive to some of the same dynamical quantities which are discussed in CQ theories. Finally, I conducted two experiments specifically to address the interaction of CQ theories and causal perception. In addition, in passing I pointed to neurobiological mechanisms which show promise of explaining how our perceptions actually do latch onto causal properties through preprocessing in the motion detection system. At the least these findings demonstrate that the correct ontology for causation is one of events *extended* in time and space.

Finally, I have demonstrated the last point by example. The fact that empirical science can help undermine the very perceptual ontology which got Hume into his bind in the first place is one recommendation for this approach. The fact that causality is both an objective feature of the world and one whose characteristics are strongly constrained by the role it plays in human perception and understanding of the world is another. The fact that CQ theory gains support for its relevance to everyday notions of causality through a grounding in the results of event perception is another. And last, the ability to formulate CQ theory in such a way as to afford testing of its claims as I did in the final chapter offers a way to move philosophy out of the realm of pure intuition and toy examples. It does not transform philosophy into empirical science, because the questions of philosophy ultimately go beyond what empirical science can answer. Neither, however, is worthwhile philosophy about empirical concepts like causality and perception going to get off the ground without paying attention to some empirical information, and perhaps occasionally adding some empirical methods to the tool box.

The philosophy of science is concerned, among other things, with the nature of science, its relation to the world, and its role in the human endeavor. As I said at the outset, I take it as given that the purpose of science is to yield understanding of the natural world. Science achieves that purpose largely by explaining how the world works, and the explanations given by science are often—but not always: see Friedman [37, 38] and Berger [4]—causal in nature. I suspect there is more than a coincidental relationship with the findings that people prefer mechanism-style explanations of everyday events to other alternatives [2, 1]. Causation is fundamental to the purpose and practice of science.

This perceptual account I have developed here will not help science discover causal links in cases like smoking and lung cancer, or ice-cream sales and crime rates, but it does ground the nature of the causal relation which those endeavors seek to discover without the possibility of direct perception. The discovery of causes in science has a strict analogy with the kinds of conscious causal inference research being pursued by Patricia Cheng and Barbara Spellman (see [17, 16]). One thing common to all those accounts is that they have to assume some background set of implicit causal knowledge—usually formulated as a stock idea of causal mechanisms—about which the statistical inference theory being developed can say nothing.

The fact that the hybrid CQ-perception theory I have elaborated can give a single unified account of both perceptual causation and at the same time some ultimate reduction of the idea of causation is intrinsically interesting in that it does seem to help fill in the think hole left by Hume's destructive skepticism. However, I also believe that philosophy of science should endeavor to be of some conceptual assistance to science, at the very least by keeping track of good methodologies. Towards this end I see the ultimate value of this work, if it stands, as helping to provide a foundation for broader accounts of causal inference and causal discovery such as those set within the framework of causal modeling. Once again, we are faced with the parallels between pursuing philosophy and psychology. At least some of the inference methods sought by those conducting research on causal discovery are already employed by observers every day in the task of getting around the world successfully.

Hume's references to Newton

James Force [34] has compiled a list of references to Newton and Newton's works in Hume's writing.

1. *Treatise*, Appendix, pp.638-9 in 1979 printing of 2nd edition
2. "Of the Middle Station of Life," in *The Philosophical Works*, IV:379, 1882.
3. "On the Rise and Progress of the Arts and Sciences", in *Works*, III:183.
4. *A Letter from a Gentleman to his friend in Edinburgh*, 1967. pp.28-9
5. *Enquiry* 3rd ed. p.73n
6. *Enquiry Concerning the Principles of Morals*
7. Letter to cousin Mrs. Dysart 1751 (facetious)
8. *Dialogues* 1947 ed., p.136
9. *History of England*, v.VI, 1782 ed. 196-7.
10. *HoE* VIII:332-4
11. *Nat'l Hist. of Religion*1976, p.79

Chronology for Newton & Hume

1687:	*Principia*
1688:	Locke's (anonymous) review in the *Bibliothèque Universelle*
1690:	Locke's *Essay Concerning Human Understanding*
1694:	Locke's *Essay*, second edition
1704:	*Opticks*
1706:	Locke's *Essay*, fifth edition (posthumous)
1710:	Leibniz' *Theodicy*
	Berkeley's *Treatise*
1713:	*Principia*, second edition (includes General Scholium)
	Berkeley's *Three Dialogues*
1717:	Leibniz-Clarke correspondence published
1721:	Berkeley's *De motu*
1726:	*Principia*, third edition
1727:	Newton dies; Fontenelle's *Eloge*
1728:	*Chronology of Ancient Kingdoms Amended*
1729:	Motte translation
1732:	Berkeley's *Alciphron*
1733:	Newton's *Prophecies of Daniel, and the Apocalypse of St John*
1734:	Barrow's *Usefulness of Mathematical Learning*
	Hume's letter to Dr. Arbuthnot ("new scene of thought")
	Berkeley's *Treatise*, second edition
	Berkeley's *Analyst*
1738:	"Sir Isaac Newton" entry in Birch's *A General Dictionary*
1739:	Hume's *Treatise*, volumes 1 and 2
1740:	March: Hume's *Abstract*
	November: Hume's *Treatise*, volume 3
1745:	Hume's *Letter from a Gentleman to his friend in Edinburgh*
1748:	Maclaurin's *Account*
1748:	Hume's *Inquiry*, called *Philosophical Essays*
1756:	*Four Letters to Doctor Richard Bentley*
1760:	*Biographia Britannica* entry on Newton
1776:	Hume's *Dialogues Concerning Natural Religion* published

Table 1: Publication dates for relevant texts, some taken from table of biographies of Newton in [45] and the chronology of Hume in [12].

Bibliography

[1] Woo-kyoung Ahn and Jeremy Bailenson. Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology*, 31:82–123, 1995.

[2] Woo-kyoung Ahn, Charles W. Kalish, Douglas L. Medin, and Susan A. Gelman. The role of covariation versus mechanism information in causal attribution. *Cognition*, 54(3):299–352, 1995.

[3] Isaac Barrow. *The Usefulness of Mathematical Learning*. Cass Library of Science Classics. Frank Cass & Co., London, 1970/1734.

[4] Ruth Berger. *The Dynamics of Explanation*. Ph.d. thesis, Indiana University, 1997.

[5] George Berkeley. The principles of human knowledge. In David M. Armstrong, editor, *Berkeley's Philosophical Writings*, pages 41–128. Collier Books, 1965.

[6] II Berry, Michael J., Iman H. Brivanlou, Thomas A. Jordan, and Markus Meister. Anticipation of moving stimuli by the retina. *Nature*, 398:334–337, 25 March 1999. Letters to *Nature*.

[7] Geoffrey P. Bingham. Dynamical systems and event perception: A working paper, parts i-iii. *Perception/Action Workshop Review*, 2:4–7, 1987.

[8] Geoffrey P. Bingham. Dynamics and the problem of visual event recognition. In Timothy Van Gelder and Robert Port, editors, *Mind as Motion: Explorations in the Dynamics of Cognition*. Bradford Books, MIT Press, Cambridge, MA, 1995.

[9] Geoffrey P. Bingham, Richard C. Schmidt, and Lawrence D. Rosenblum. Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6):1473–1493, 1995.

[10] Myles Brand, editor. *The Nature of Causation*. University of Illinois Press, 1976.

[11] John Hill Burton. *Life and Correspondence of David Hume*, volume I. Burt Franklin, New York, 1967.

[12] Nicholas Capaldi. *David Hume: The Newtonian Philosopher*. Twayne, Boston, 1975.

[13] Nancy Cartwright. *Nature's Capacities and their Measurement*. Clarendon Press, Oxford, New York, 1989.

[14] P. Casini. Newton's *Principia* and the philosophers of the Enlightenment. In D.G. King-Hele and A.R. Hall, editors, *Newton's* Principia *and its Legacy*, volume 42 of *Notes and Records of the Royal Society*, pages 35–52. Cambridge Univ. Press, London, 1988.

[15] Pravas Jivan Chaudhury. Newton and hypothesis. *Philosophy and Phenomenological Research*, 22:344–353, 1962.

[16] Patricia W. Cheng. From covariation to causation: A causal power theory. *Psychological Review*, 104(2):367–405, 1997.

[17] Patricia W. Cheng and Keith J. Holyoak. Complex adaptive systems as intuitive statisticians: Causality, contingency, and prediction. In J.-A. Meyer and H. Roitblat, editors, *Comparative approaches to cognition*, pages 271–302. MIT Press, Cambridge, MA, 1995.

[18] John Clement. A conceptual model discussed by Galileo and used intuitively by physics students. In Dedre Gentner and Albert L. Stevens, editors, *Mental Models*, Cognitive Science, pages 325–340. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

[19] I. Bernard Cohen. History and the philosopher of science. In Frederick Suppe, editor, *The Structure of Scientific Theories*, pages 308–349. Univ. of Illinois Press, Urbana, second edition, 1979.

[20] I. Bernard Cohen. Newton's third law and universal gravitation. In P.B. Scheurer and G. DeBrock, editors, *Newton's Scientific and Philosophical Legacy*, International Archives of the History of Ideas, pages 25–53. Kluwer, Boston, 1988. Proceedings of 1987 colloquium in Nijmegen.

[21] Alan Costall. Phenomenal causality. In Thinès et al. [109], pages 51–64.

[22] Graciela DePierris. Hume's pyrrhonian skepticism and the belief in causal laws. *Journal of the History of Philosophy*, forthcoming.

[23] Phil Dowe. Process causality and asymmetry. *Erkenntnis*, 37:179–196, 1992.

[24] Phil Dowe. Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science*, 59:195–216, 1992.

[25] Phil Dowe. Causality and conserved quantities: A reply to Salmon. *Philosophy of Science*, 62:321–333, 1995.

[26] Phil Dowe. Backwards causation and the direction of causal processes. *Mind*, 105:227–248, 1996.

[27] Lisa J. Downing. Locke's Newtonianism and Lockean Newtonianism. *Perspectives on Science*, 5(3):285–310, 1997. Paper presented at the First International Conference of HOPOS, April 1996.

[28] Fred I. Dretske. Laws of nature. *Philosophy of Science*, 44:248–268, 1977.

[29] C.J. Ducasse. Causality: Critique of Hume's analysis. In Myles Brand, editor, *The Nature of Causation*, pages 67–76. University of Illinois, 1976. From C.J. Ducasse, *Nature, Mind and Death*, 1951.

[30] C.J. Ducasse. On the nature of the observability of the causal relation. In Ernest Sosa and Michael Tooley, editors, *Causation*, Oxford Readings in Philosophy, chapter VII, pages 125–136. Oxford, New York, 1993. Originally *Journal of Philosophy*, 23(1926), 57-68.

[31] David Fair. Causation and the flow of energy. *Erkenntnis*, 14:219–250, 1979.

[32] Marta Fehèr. The method of analysis-synthesis and the structure of causal explanation in Newton. *International Studies in the Philosophy of Science*, 1(1):60–84, 1986.

[33] Mordechai Feingold. Partnership in glory: Newton and Locke through the Enlightenment and beyond. In p.B. Scheurer and G. Debrock, editors, *Newton's Scientific and Philosophical Legacy*, pages 291–308. Kluwer, 1988.

[34] James Force. Hume's interest in Newton and science. *Hume Studies*, XIII(2):166–217, 1987.

[35] Marina Frasca-Spada. Some features of Hume's conception of space. *SHPS*, 21(3):371–411, 1990.

[36] Marina Frasca-Spada. *Space and the Self in Hume's* Treatise. Cambridge University Press, New York, 1998.

[37] Michael Friedman. Explanation and scientific understanding. LXXI(1):5–19, 1974.

[38] Michael Friedman. Theoretical explanation. In R. Healey, editor, *Reduction, Time, and Reality*, pages 1–16. Cambridge Univ. Press, Cambridge, 1981.

[39] Alan Gabbey. Huygens and mechanics. In *et al* H.J.M. Bos, editor, *Studies on Christian Huygens*, pages 166–199. Lisse, 1980.

[40] Karl Gegenfurtner. The eyes have it! *Nature*, 398:291–292, 25 March 1999.

[41] David L. Gilden. Heuristic judgment of mass ratio in two-body collisions. *Perception & Psychophysics*, 56(6):708–720, 1994.

[42] David L. Gilden and Dennis R. Proffitt. Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):372–383, 1989.

[43] Clark Glymour and Patricia Cheng. Causal mechanism and probability: a normative approach. In M. Oaksford and N. Chater, editors, *Rational models of cognition*. Oxford Univ. Press, Oxford, 1999. IN PRESS.

[44] Alvin I. Goldman. A causal theory of knowledge. *Journal of Philosophy*, 64(12):357–372, 1967.

[45] A. Rupert Hall. *Isaac Newton: Eighteenth-century Perspectives*. Oxford, New York, 1999.

[46] Jr. Harvey, Lewis O. *Rscore+ v.5.1*, July 1999.

[47] Ahmad I., M.D. Hashem and Gregory F., M.D., Ph.D. Cooper. Human causal discovery from observational data. In *Proceedings of the 1996 symposium of the American Medical Information Association*, pages 27–31. American Medical Information Association, 1996.

[48] Leslie A. Hayduk. *Structural Equation Modeling with LISREL: Essentials and Advances*. Johns Hopkins, Baltimore, 1987.

[49] Patrick J. Hayes. The second naive physics manifesto. In Jerry R. Hobbs and Robert C. Moore, editors, *Formal Theories of the Commonsense World*, Ablex series in Artificial Intelligence, pages 1–36. Ablex Publishing Corp., Menlo Park, 1985.

[50] Heiko Hecht. Heuristics and invariants in dynamic event perception: Immunized concepts or nonstatements? *Psychonomic Bulletin & Review*, 3(1):61–70, 1996.

[51] Heiko Hecht, Mary K. Kaiser, and Martin S. Banks. Gravitational acceleration as a cue for absolute size and distance? *Perception & Psychophysics*, 58(7):1066–1075, 1996.

[52] Heiko Hecht and Dennis Proffitt. What cues do billiard experts use? Conceptual and perceptual judgments of spin. Manuscript, 1994ish.

[53] David Hume. *A Treatise of Human Nature*. Oxford/Clarendon, London, 1739-40. 1973 printing of Selby-Bigge's 1888 edition.

[54] David Hume. An abstract of a treatise of human nature. In Charles W. Hendel, editor, *An Inquiry Conerning Human Understanding*, volume 107 of *The Library of Liberal Arts*, appendix II, pages 183–198. Bobbs-Merrill, Indianapolis, 1955 edition, 1740/1955. Supplement to edition of *Inquiry*.

[55] David Hume. *An Inquiry Concerning Human Understanding*. Bobbs-Merrill, Indianapolis, 1955 edition, 1748/1955.

[56] David Hume. *Dialogues Concerning Natural Religion*. Hackett, 1776/1988.

[57] David Hume. *An Enquiry Concerning the Principles of Morals*. Open Court, La Salle, Illinois, 1966. Originally 1751. Reprinted from the 1777 edition.

[58] Paul Humphreys. *The Chances of Explanation*. Princeton University Press, 1989.

[59] Gurol Irzik. Causal modeling and the statistical analysis of causation. In *PSA 1986*, volume 1, pages 12–23, 1986.

[60] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2):201–211, 1973.

[61] Gunnar Johansson, Claes von Hofsten, and Gunnar Jansson. Event perception. *Ann. Rev. Psychol.*, 31:27–63, 1980.

[62] Mary Kister Kaiser and Dennis R. Proffitt. Observers' sensitivity to dynamic anomalies in collisions. *Perception & Psychophysics*, 42(3):275–280, 1987.

[63] Jaegwon Kim. Causes and events: Mackie on causation. In Ernest Sosa and Michael Tooley, editors, *Causation*, pages 60–74. Oxford Univ. Press, New York, 1993.

[64] Knowledge Revolution, San Francisco. *Interactive Physics II*, 1992.

[65] Knowledge Revolution, San Francisco. *Interactive Physics 5*, 1998.

[66] Kevin B. Korb. The frame problem: An ai fairy story. 1997.

[67] Kevin B. Korb. Probabilistic causal structure. In H. Sankey, editor, *Causation and Laws of Nature*. Kluwer, 1999. Forthcoming.

[68] John Kruschke. The perception of causality: A performance measure of ampliation, 16 May 1987 1987.

[69] John K. Kruschke and Michael M. Fragassi. The perception of causality: Feature binding in interacting objects. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, pages 441–446. Cognitive Science Society, Erlbaum, 1996.

[70] Alan M. Leslie. Spatiotemporal continuity and the perception of causality in infants. *Perception*, 13:287–305, 1984.

[71] Alan M. Leslie and Stephanie Keeble. Do six-month-old infants perceive causality? *Cognition*, 25:265–288, 1987.

[72] David Lewis. Causation. In Ernest Sosa and Michael Tooley, editors, *Causation*, pages 193–204. Oxford Univ. Press, New York, 1993.

[73] John Locke. *An Essay Concerning Human Understanding*. Meridian British Philosophers Series. Meridian Books, New York, abridged and edited 1964 edition, 1690. Introduction by A.D. Woozley.

[74] J.L. Mackie. *The Cement of the Universe*. Oxford, 1974.

[75] Colin Maclaurin. *An account of Sir Isaac Newton's philosophical discoveries*. Patrick Murdoch, 1748. 392p., Imprimatur Folkes [Pres.Roy.Soc.] at Royal Society meeting June 9, 1748. Dedication to the late author by his wife Anne Maclaurin. Also, a "Life of the Author".

[76] Neil A. MacMillan and C. Douglas Creelman. *Detection theory: a user's guide*. Cambridge Univ. Press, 1991.

[77] MacroMind, San Francisco. *MacroMind Accelerator*, 1989. Computer Software.

[78] Michael McCloskey. Naive theories of motion. In Dedre Gentner and Albert L. Stevens, editors, *Mental Models*, Cognitive Science, pages 299–324. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

[79] Daniel S. McConnell, Michael M. Muchisky, and Geoffrey P. Bingham. The use of time and trajectory forms as visual information about spatial scale in events. *Perception & Psychophysics*, 60:1175–1187, 1998.

[80] Albert Eduard Michotte. *The Perception of Causality*. Methuen & Co. Ltd., London, 1963. ed. by C.A. Mace. Originally published under *La Perception de la Causalité,* 1946 and 1954. BF467.M62.

[81] T.R. Miles. *Commentary*, chapter Commentary, pages 373–415. Basic Books, 1963. C.A. Mace, ed. Summary of Michotte and essays.

[82] Michael M. Muchisky and Geoffrey P. Bingham. Perceiving size in events via kinematic form. In John Kruschke, editor, *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, pages 1002–1007. Erlbaum, 1992.

[83] Ernest Nagel. The logical character of scientific laws. In Myles Brand, editor, *The Nature of Causation*, pages 77–110. University of Illinois, 1976. From *The Structure of Science*, 1961.

[84] Isaac Newton. *Sir Isaac Newton's Mathematical Principles of Natural Philosophy*. Cambridge, 1947. Translated by Andrew Motte and revised by F. Cajori.

[85] Isaac Newton. *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the World*. University of California Press, Berkeley, 1960. Trans. Andrew Motte, 1729. Ed. Florian Cajori. Fourth printing.

[86] David Fate Norton and Mary J. Norton. *The David Hume Library*. Edinburgh Bibliographical Society, Aberystwyth, 1996.

[87] James Noxon. *Hume's Philosophical Development*. Clarendon, Oxford, 1973.

[88] Lisa M. Oakes. Development of infants' use of continuity cues in their perception of causality. *Developmental Psychology*, 30(6):869–879, 1994.

[89] John B. Pittenger. Estimation of pendulum length from information in motion. *Perception*, 14:247–256, 1985.

[90] John B. Pittenger. Detection of violations of the law of pendulum motion: observers' sensitivity to the relation between period and length. *Ecological Psychology*, 2(1):55–71, 1990.

[91] Dennis R. Proffitt and David L. Gilden. Understanding natural dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):384–393, 1989.

[92] Dennis R. Proffitt, Mary K. Kaiser, and Susan M. Whelan. Understanding wheel dynamics. *Cognitive Psychology*, 22:342–373, 1990.

[93] G.A.J. Rogers. Locke's essay and Newton's Principia. *Journal of the History of Ideas*, 39(2):217–232, 1978.

[94] Sverker Runeson. *On Visual Perception of Dynamic Events*. PhD thesis, Uppsala, 1977.

[95] Sverker Runeson. Support for the cue-heuristic model is based on suboptimal observer performance: Response to Gilden and Proffitt (1994). *Perception & Psychophysics*, 57(8):1262–1273, 1995.

[96] Sverker Runeson and Dankert Vedeler. The indispensability of precollision kinematics in the visual perception of relative mass. *Perception & Psychophysics*, 53(6):617–632, 1993.

[97] Wesley C. Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton, 1984.

[98] Wesley C. Salmon. Causality without counterfactuals. *Philosophy of Science*, 61:297–312, 1994.

[99] Wesley C. Salmon. Causality and explanation: A reply to two critiques. *Philosophy of Science*, 64(3):461–477, 1997.

[100] Bror V.H. Saxberg. Projected free fall trajectories I: Theory and simulation. *Biological Cybernetics*, 56:159–175, 1987. Article 1 of 2.

[101] Bror V.H. Saxberg. Projected free fall trajectories II: Human experiments. *Biological Cybernetics*, 56:177–184, 1987. Article 2 of 2.

[102] Christine M. Shepherd. Newtonianism in Scottish universities in the seventeenth century. In R.H. Campbell and Andrew S. Skinner, editors, *The Origins and Nature of the Scottish Enlightenment*, pages 65–85. John Donald Publishers, Ltd., Edinburgh, 1982.

[103] Norman Kemp Smith. *The Philosophy of David Hume*. Macmillan, New York, 1960.

[104] Ernest Sosa and Michael Tooley, editors. *Causation*. Oxford Readings in the Philosophy of Science. Oxford Univ. Press, New York, 1993.

[105] Barbara A. Spellman. Acting as intuitive scientists: Contingency judgments are made while controlling for alternative potential causes. *Psychological Science*, 7:337–342, 1996.

[106] Pieter Jan Stappers and Patrick E. Waller. Using the free fall of objects under gravity for visual depth estimation. *Bulletin of the Psychonomic Society*, 31(2):125–127, 1993. Paul J. Locher, *sponsor*.

[107] Barry Stroud. *Hume*. Arguments of the Philosophers. Routledge, New York, 1977. Ted Honderich, ed.

[108] H.S. Thayer. *Newton's Philosophy of Nature: selections from his writings*. New York, 1953.

[109] Georges Thinès, Alan Costall, and George Butterworth, editors. *Michotte's Experimental Phenomenology of Perception*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.

[110] J. Todd and W. Warren. Visual perception of relative mass in dynamic events. *Perception*, 11:325–335, 1982.

[111] H.W. Turnbull, editor. *The Correspondence of Isaac Newton*, volume 1. Cambridge Univ. Press, 1959.

[112] William A. Wallace. *Causality and Scientific Explanation*, volume 1. Univ. of Michigan Press, Ann Arbor, 1974. Medieval and Early Classical Science.

[113] William A. Wallace. *Causality and Scientific Explanation*, volume 2. Univ. of Michigan Press, Ann Arbor, 1974. Classical and Contemporary Science.

[114] Sylvia Weir. The perception of motion: Michotte revisited. *Perception*, 7:247–260, 1978.

[115] Richard S. Westfall. *Force in Newton's Physics*. History of Science Library. American Elsevier, 1971.

[116] Richard S. Westfall. *Never at Rest: A Biography of Isaac Newton.* Cambridge University Press, 1980. 1996 ed.

[117] Emily A. Wickelgren and Geoffrey P. Bingham. Infant sensitivity to trajectory form. *Journal of Experimental Psychology: Human Perception and Performance*, Submitted.

[118] Jr. William H. Warren and Robert E. Shaw, editors. *About Visual Event Perception*, Hillsdale, New Jersey, 1985. Lawrence Erlbaum Associates. Proceedings of the First International Conference on Event Perception.