

Continuing Commentary

Commentary on P. J. H. Schoemaker (1991). The quest for optimality: A positive heuristic of science? *BBS* 14(2):205–215.

Abstract of the original target article: This paper examines the strengths and weaknesses of one of science's most pervasive and flexible metaprinciples; *optimality* is used to explain utility maximization in economics, least effort principles in physics, entropy in chemistry, and survival of the fittest in biology. Fermat's principle of least time involves both teleological and causal considerations, two distinct modes of explanation resting on poorly understood psychological perspectives. The rationality heuristic in economics provides an example from social science of the potential biases arising from the extreme flexibility of optimality considerations, including selective search for confirming evidence, ex post rationalization, and the confusion of prediction with explanation. Commentators are asked to reflect on the extent to which optimality is (1) an organizing principle of nature, (2) a set of relatively unconnected techniques of science, (3) a normative principle for rational choice and social organization, (4) a metaphysical way of looking at the world, or (5) something else still.

Huygens' Principle: A case against optimality

Hans-Martin Gaertner

Zentrum für Allgemeine Sprachwissenschaft (ZAS) Berlin, D-10117 Berlin, Germany. gaertner@DMZ01.zas.gwz-berlin.de

Abstract: I will present evidence that nature does not optimize in the sense of Fermat's principle of least time, contrary to what Schoemaker's unintentionally ambiguous exposition might suggest. First, Huygens' principle, an alternative nonteleological account of Snell's law, is outlined. Second, I confront Fermat's principle with a substantive conceptual problem.

Schoemaker (1991) uses Fermat's principle of least time, which derives Snell's law of the refraction of light, to shed light on the difference between teleological and causal theories. Indeed, to focus on the science of physics in exploring metascientific principles, such as optimality, is a plausible strategy, given that understanding in the area of physics has proceeded to a point where more than plausibility can be appealed to. It is all the more important, however, to be as clear as possible in the interpretation of the results obtained. In that respect, Schoemaker's exposition of Fermat's principle – if we accept for the sake of argument that extremum and optimality principles can be fruitfully examined in parallel – is likely to generate misunderstanding as to the “reality” of optimality.

Although Schoemaker grants that “physics, the most sophisticated of the optimality sciences, has the fewest *a priori* arguments in its [optimality's] favor” (p. 208), who would care, as long as such principles help us achieve “excellent numerical results for otherwise intractable problems” (Schoemaker 1991, p. 206)? (The latter remark was not directly applied to Fermat's principle but to Hamilton's law, which is taken to be a generalization of Fermat's principle among others [p. 206]). (Cf. Bookstein's [1991, pp. 216ff] commentary on Hamilton's law.)

Historically, however, Fermat's principle has been rivaled by Huygens's wave theory of light, which in 1678, only 13 years after Fermat's death, likewise derived Snell's law.

Huygens' Principle (for the following, cf. Sears et al. 1976) is a geometrical method for constructing the shape of a wave front at t' from the shape of a wave front at an earlier point t . Consider Figure 1 on the following page.

At t , AA' is our wave front. Crucially, any point on AA' can be taken to give rise to a secondary wavelet spreading in all directions. Thus, infinitely many circles (of radius n) can be drawn off AA' to reach the new wave front BB' . BB' is the surface tangent to the secondary wavelets, called *envelope*. How to interpret the inner portions of the wavelets need not concern us here.

Snell's law of the refraction of light follows. Consider Figure 2.

X is the surface between two transparent media, a and b . AA' is our incoming wave front of light. A' is chosen so that the circle of the secondary wavelet of magnitude v_1t emanating from A' will cut the surface at B' , AA' , and $A'B'$, forming a right angle. A is a point on X , and B lies at a distance of v_2t from A . We derive the relation between Θ_1 , the angle of incidence, and Θ_2 , the angle of refraction, via the respective sinus functions:

$$\sin\Theta_1 = v_1t/AB', \text{ and } \sin\Theta_2 = v_2t/AB', \text{ which gives us}$$

$$v_1t/\sin\Theta_1 = v_2t/\sin\Theta_2, \text{ hence}$$

$$\sin\Theta_1/\sin\Theta_2 = v_1/v_2 \text{ or } \sin\Theta_1 = v_1\sin\Theta_2/v_2$$

that is, the formulation of Fermat's principle (Schoemaker 1991, p. 209).

Obviously, Huygens' principle is not an optimality principle. The concept of least time is not invoked. Therefore, my first conclusion is that whatever we consider to be a plausible theory of light, we are not justified in taking the “intractability” of Snell's law to favor Fermat's principle.

A second area of potential misunderstanding is Feynman et al.'s (1964) discussion of Fermat's principle, which Schoemaker adduces. Thus,

light decides which is the shortest time, or the extreme one and chooses the path. But *what* does it do, *how* does it find out? Does it *smell* the nearby paths, and check them against each other? The answer is, yes, it does, in a way. (Schoemaker 1991, p. 209)

Here, I would like to concentrate on what it means that quantum theory gives “considerable justification” (Schoemaker 1991, p. 209) to Fermat's principle. “Fermat was ingenious to have formulated a principle that accords so well with a deeper reality that was unrecognized at the time” (p. 209). A weak interpretation of *justification*, the one intended, I think, by Schoemaker, would be that Fermat's principle provided a useful heuristic. However, Feynman's further remarks, referring to “people who do not like

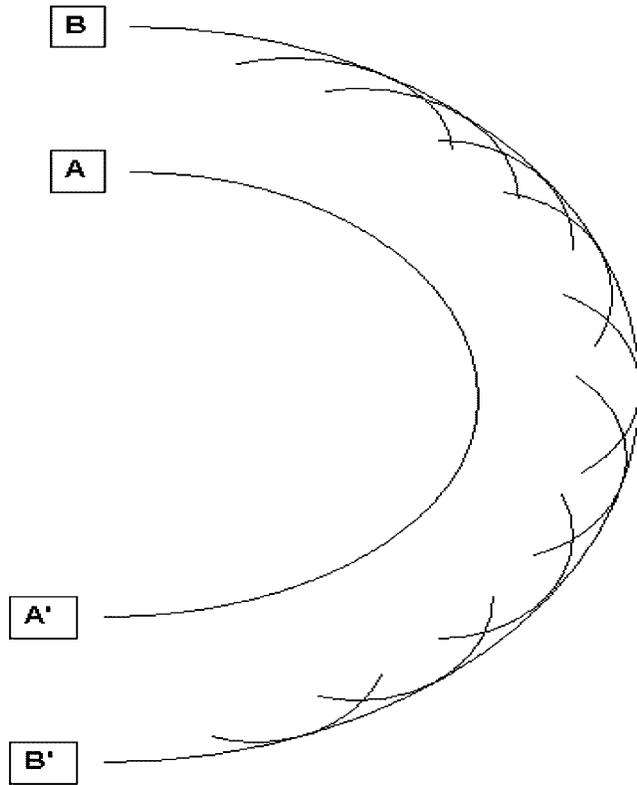


Figure 1 (Gaertner). Wavelet theory of the propagation of light from wave front AA' at t to BB' at t' .

this kind of theory" (i.e., the principle of least time) (Schoemaker 1991, p. 209) drives at a stronger, "realistic" interpretation. Again, however, Huygens' principle seems to be equally justified by the photon view of light refraction advocated by Feynman. Thus, reinforcing and canceling vectors accord excellently with Huygens's secondary wavelets (pace Feynman et al. 1964, Ch. 26, p. 8).

Of course, this issue is complicated, given the well-known wave/particle dichotomy attributed to light. My point is that no strong conclusions should be drawn concerning optimality. Note also that Schoemaker's discussion is based on Feynman et al.'s Chapter 26, which "is one of the chapters that we shall have to 'unlearn' again" (1964, p. 1).

Historically, Huygens had to face the opposition of Newton's corpuscular theory of light (cf. Feyerabend 1981; Gamow 1961). During the nineteenth century, it was shown to be superior to a corpuscular theory in the area of *interference* and *diffraction* by Thomas Young, Augustin Fresnel, and Leon Foucault (cf. Sears et al. 1976, p. 639), and was propped up by Maxwell's theory of electromagnetic waves. In a philosophically unpretentious way, Sears et al. conclude that "the phenomena of light *propagation* may best be described by the electromagnetic *wave* theory" (1976, p. 640).

If we take Feynman's lead, however, we could point to the fact that "both teleological and causal considerations . . . rest on poorly understood psychological primitives" (Schoemaker 1991, p. 205). Although Huygens' principle appears to provide the "process explanation" that, according to Schoemaker, "each optimality principle [Fermat's principle in our case] begs for" (p. 212), preferring Huygens's principle on these grounds might be just our psychological bias.

To counter this argument, I would finally like to present a conceptual problem for Fermat's principle having to do with the *reflection* of light. Consider Figure 3 on the following page.

X is a reflecting surface. We know that $\Theta_1 = \Theta_2$. This can be derived from Huygens's principle in a straightforward way (Sears et al. 1976, pp. 654f). The concept of least time taken literally, however, results in a puzzle. If light *wants* to get from (point) A to

(point) B, the path of least time within an isotropic medium would be a straight line. If the principle of least time is an absolute property of the nature of light, the phenomenon of reflection is predicted not to exist, contrafactually, of course.

Huygens' principle has the clear advantage of presenting a unified account of the refraction *and* the reflection of light. Fermat's optimality principle as such is not sufficient to account for both. The factor that makes one feel uneasy in the case of the refraction of light turns into a real problem when it comes to the analysis of the reflection of light. Explicit boundary conditions have to be given to guarantee that the principle of least time applies only in the right context.

What is the way to get from A to B in the shortest time? The answer is to go straight from A to B! But *if we add the extra rule that the light has to strike the mirror and come back in the shortest time*, the answer is not so easy. (Feynman et al. 1964, Ch. 26, p. 3; emphasis mine)

Adding such extra rules artificially glosses over what is problematic about Fermat's principle; namely, that, for it to apply non-trivially, it has to violate the Nahewirkungsprinzip (action at short distance) reintroduced by Einstein's theory of relativity (cf. von Weizsaecker 1994, p. 236).

In summary, I argue that in the case of Fermat's principle of least time there are clear advantages to saying that nature does not optimize.

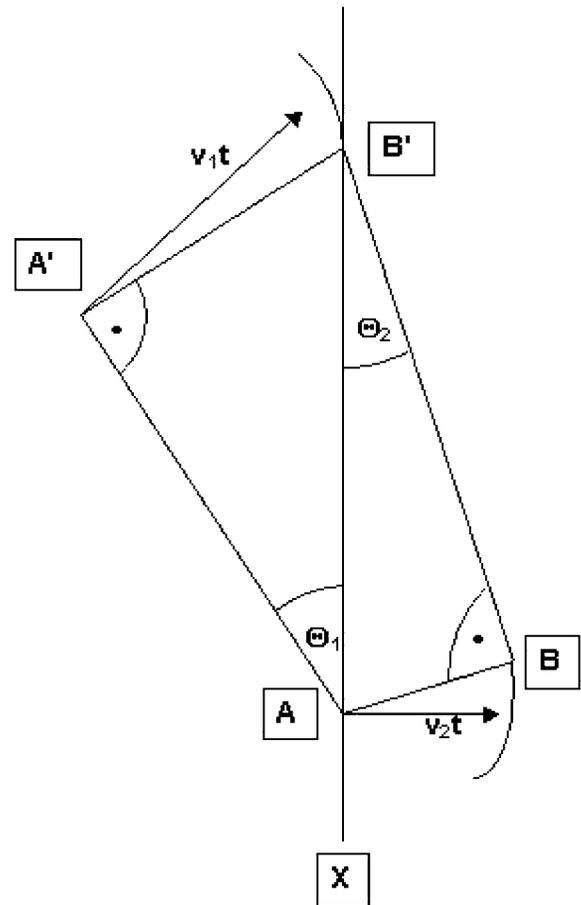


Figure 2 (Gaertner). Refraction of incoming wave front of light AA' at surface X, resulting in refracted wave front BB'.

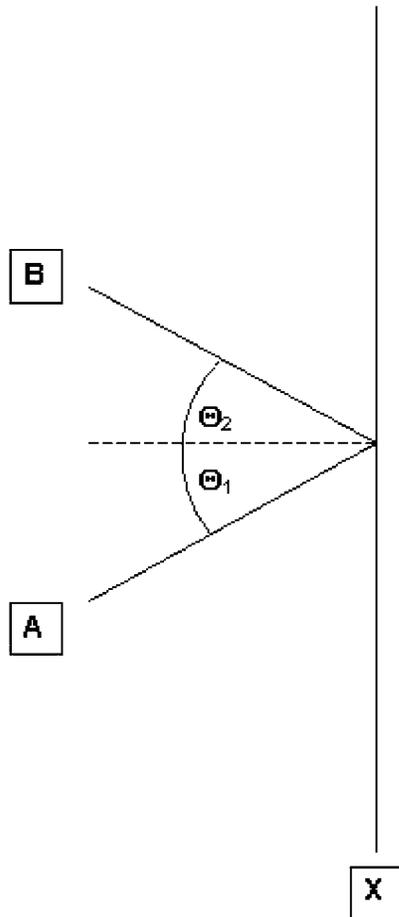


Figure 3 (Gaertner). Reflection of (a single beam of) light emanating from A at surface X.

ACKNOWLEDGMENTS

This work was written while I was holding a grant from the Max-Planck-Gesellschaft (ASG, Berlin). Thanks for comments and inspiration go to Reinhard Blutner and Anatoli Strigin.

Author's Response

Huygens versus Fermat: No clear winner

Paul J. H. Schoemaker

Department of Marketing, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104. schoemak@wharton.upenn.edu

Abstract: How should we assess the appeal of multiple scientific theories when they can all explain a particular empirical phenomenon of interest? We contrast Huygens' and Fermat's explanations of the law of refraction of light and find that neither dominates the other when considering multiple criteria for assessing the overall appeal of a scientific theory. The absence of teleology in Huygens' account is a strong plus compared to Fermat's. But Huygens' wave theory scores less well with respect to other desiderata for a scientific theory. In this case, there does not appear to be a clear winner, nor need there be one.

Gaertner provides valuable insights into my original discussion of Fermat's principle of least time (Schoemaker 1991). He reminds us – as well as demonstrates – that Huygens' wave theory can also be used to derive Snell's law of refraction but without resort to the metaphysical notion that light travels the shortest distance in time. He considers this “a case against optimality.”

If we desire a purely causal or descriptive account of nature's laws, Huygens will be much more appealing than Fermat. Huygens clearly offers an alternative account of Snell's law, but whether this constitutes a case against optimality is debatable. For example, if parsimony or simplicity is important, Fermat's account may be preferred. It is far simpler to explain – to, say, high school students – than Huygens' interference pattern. The latter entails the cumulative effect of waves that are slightly or greatly out of phase.

Further, suppose we judge a scientific theory not only on its ability to explain observed empirical phenomena *ex post facto*, but also on its ability to predict new phenomena *ex ante*. In that case, Fermat may have the upper hand. The principle of least time is deceptively simple, both to apply and to remember. And the deception lies in its appeal to teleological explanations. But as long as the teleological approach results in valid new predictions – as Fermat's principle did about the behavior of light in converging lenses – its metaphysical nature may well be a price worth paying. The key is to view the search for optimality as a powerful heuristic. This may sound paradoxical since heuristics are by definition not optimal. But we should not take this perceived optimality too literally, as though it were a deep scientific truth. At this stage, we simply don't know.

R1. Multiple criteria

It would have been interesting if Gaertner could have scored Fermat versus Huygens on these multiple criteria, rather than place all his weight on one (the absence of teleology). This perhaps underscores the subjectivity of science. What constitutes an adequate scientific explanation? Clearly, explaining the empirical phenomenon at hand ranks very high. But I think that simplicity and elegance, as well as the propensity to spawn new predictions, constitute important criteria, too. And consistency with other prevailing scientific explanations, or the breadth of the domain of application, should perhaps matter as well. Table R1 summarizes some of the criteria I would personally use to evaluate competing theories.

Table R1. Evaluating a Scientific Theory

Possible Criteria to Use:

- a. Does it explain the phenomenon well?
- b. Is the explanation easy to follow or apply?
- c. Is it a parsimonious explanation (fewest assumptions)?
- d. Does it generate interesting new predictions?
- e. Can the theory be falsified in principle?
- f. How consistent is it with other theories?
- g. How broad is the potential domain of application?
- h. How widely is it used by practicing scientists?
- i. Does it have strong competitors (i.e., alternative theories)?
- j. Has it been widely tested, with positive results?

Because science is a cumulative process, where better theories replace inferior ones in the Darwinian struggle for intellectual survival, it is interesting that Fermat's principle is still taught in physics courses at both the high school and university levels. And the twin existence of wave and corpuscular theories of light, more generally, underscores that neither one dominates the other on all accounts. In some applications, such as simple optics, we may prefer Fermat. In others, involving interference patterns or the phenomenon of reflection, we would presumably favor Huygens.

R3. Reflection of light

Gaertner justly notes that the application of Fermat's principle may require the specification of clear boundary conditions before it can be usefully invoked. The example Gaertner gives about light reflecting off a glass plate (see his Fig. 3) illustrates the need for boundary conditions well. The shortest path from *A* to *B* is indeed a straight line. But once the boundary condition has been imposed, that the light beam must strike the mirror en route, Fermat's principle makes the right prediction. Any other path from *A* to *B* along the mirror would be longer than the symmetric triangle drawn in that figure. And indeed, if points *A* and *B* had not been equidistant from the mirror, as in my Figure R1 here, Fermat's principle would allow for an easy solution. To find the shortest path, we can set up an algebraic distance equation to prove that *C* is indeed the point of shortest distance (and hence, time, in this isotropic medium). A simpler approach, however, is geometric analysis. If we simply reflect point *B* through the mirror to the other side (called *B**) and then draw the straight line *A* to *B**, point *C* emerges as the shortest route. Any other point beside *C*, such as, for example, *C'*, will make route *A-C'-B** longer than *A-C-B**. And because *A-C-B* is equal in distance to *A-C-B**, we have proved that *C* is the reflection point of shortest distance and time.

This visual example illustrates the intuitive simplicity and ease of application of Fermat's principle. But yes, it does require that some clear boundary conditions are set beforehand. However, all theories will require boundary conditions, and it is not so evident to me that Fermat's is more restrictive or arbitrary in this regard than Huygens'. Indeed, I wish that Gaertner had also indicated the specific boundary conditions needed for wave theory's account. And I would challenge him to offer a wave theory explanation as

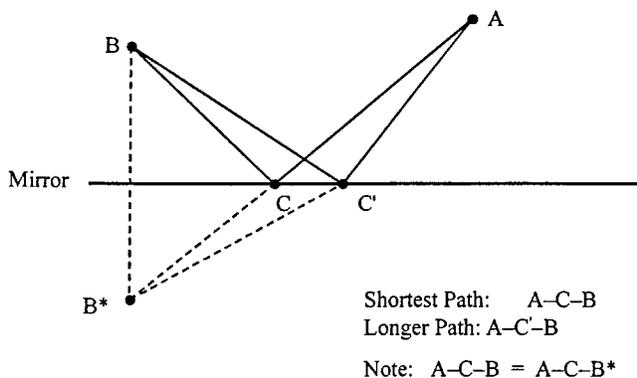


Figure R1. Reflection of light

simple as the example here, as to why point *C* will indeed be the point of reflection when a beam of light is directed from *A* to *B* via the mirror (assuming the general case where *A* and *B* are not equidistant from the mirror). To me, the beauty of Fermat's principle lies in its simplicity.

R4. Beyond physics

Of course, Snell's law is just one example – the one I happen to focus on in my target article – to make a broader statement about the pros and cons of optimality arguments. In my rejoinder to commentators in 1991, I addressed an example from the social sciences, namely, the economist's view that humans are utility-maximizing creatures. In this case, the power of a simple optimality model is quite evident. Simple explanations can be constructed and clear predictions made, about the economic affairs of organizations as well as individuals. It would require much heavier machinery to derive these same conclusions from a deeper process model that acknowledges the heuristic nature and bounded rationality of the human actor. And even psychologists, who generally reject the optimality assumptions of economics, have used optimality arguments in their descriptive models (e.g., prospect theory in risky choice, signal detection models, linguistics, etc.).

Similarly, survival of the fittest offers biologists an easy principle via which to explain the diverse and complex behavior of species. Indeed, the principle is so easy to apply that it has been criticized as too facile in hindsight, while offering limited predictive power in going forward. But the principle's basic premises (of random mutation and natural selection) are undeniable and have been demonstrated in laboratory experiments with fruit flies and the like. The challenge comes when we are applying a broad principle, such as survival of the fittest or utility maximization in economics, without undue plasticity or excessive hindsight bias. More specific principles, based on narrow causal mechanisms, may do better in this regard but also may lack the metaphoric qualities of the broader principles in sparking human imagination. This may be the fundamental trade-off we have to face when judging competing scientific accounts of observed phenomena in nature.

R5. In conclusion

I appreciate Gaertner's view that wave theory can explain Snell's law as well as Fermat's, without having to resort to metaphysics. However, I am not persuaded that Fermat's principle (and the implicit argument of optimality) is inferior to Huygens on all other counts. The need for simplicity (of explanation and application), parsimony in assumptions (beyond the teleological one) and the ability to stimulate the mind to new hypotheses are important criteria as well. Fermat's principle scores rather well on these criteria, and this was one of the prime reasons I profiled this particular principle in my 1991 target article. So the basic question remains as to why optimality arguments – which seem so deeply flawed in their explicit assumption that nature optimizes – are so effective in explaining observed phenomena in nature and, even more puzzling, in predicting entirely new ones. We are getting closer to an answer, thanks to Gaertner as well as many other commentators, but I am not sure we have a really good account yet.

References

Letter “r” appearing before author’s initials refers to Continuing Commentary Response references

- Bookstein, F. L. (1991) Optimality as a mathematical rhetoric for zeroes. Commentary on Schoemaker (1991). *Behavioral and Brain Sciences* 14: 216–17. [H-MG]
- Feyerabend, P. K. (1981) *Problems of empiricism. Philosophical papers, vol. 2.* Cambridge University Press. [H-MG]
- Feynman, R. P., Leighton, R. B. & Sands, M. (1964) *The Feynman lectures of physics, vol. I.* Addison-Wesley. [H-MG]
- Gamow, G. (1961) *The great physicists from Galileo to Einstein.* Dover. [H-MG]
- Schoemaker, P. J. H. (1991) The quest for optimality: A positive heuristic of science? *Behavioral and Brain Sciences* 14:205–45. [H-MG, rP]HS]
- Sears, F., Zemansky, M. & Young, H. (1976) *University physics, part II.* Addison-Wesley. [H-MG]
- Von Weizsaecker, C. F. (1994) *Aufbau der physik.* Deutscher Taschenbuch Verlag. [H-MG]