Teaching Physics: Inquiry and the Ray Model of Light

Fernand Brunschwig, M.A.T. Program, Hudson Valley Center

My thinking about these matters was stimulated by my participation on a panel devoted to "Inquiry in Physics Teaching" organized by Don S. Cook of Bank Street College and Fran Wald of New York University at the New York Academy of Sciences in March 2003. The specific sequence of questions as a way to provoke inquiry in physics was originally developed by Dewey Dykstra of Boise State University.

The popular conception of physics involves solving equations to obtain accurate numerical results, which correspond closely with the results of experiments. But actually learning (or teaching) physics involves mastering how to use a scientific model – an ideal mental representation underlying the observed phenomena. The ability to use (that is, manipulate mentally) a theoretical model is what is referred to as "understanding," and this is what gives rise to the equations themselves, as well as the ability to solve problems and to predict what will happen in a given situation.

The phenomena of light (particularly images formed by pinholes and lenses) offers a good example of the issues involved in learning how to use a scientific model. A good scientific model for this situation is called the ray model of light. The basic feature of this model is that light is assumed to travel in straight lines, called rays, from sources (for example a glowing object) to receivers (for example, a screen or our eyes). If a student is to use this model successfully, he or she must, to the extent possible, engage in inquiry – that is, constructing and testing his or her own mental version of the ray model, and applying it independently in a variety of situations.

An aside on approaches. A teacher can supply a detailed list of the assumptions of the ray model, and I will summarize these later, but I believe, and most scientists agree, that in order to be able to use a model successfully it is not really useful to simply memorize and recall a static list of assumptions. Rather, one must build up a more comprehensive mental picture of the model from which the assumptions can be derived, or into which the assumptions are integrated. Students differ in the way they assimilate information, and fundamentalists could argue that memorizing the assumptions is the obvious first step. I won't disagree that this might be the best first step for some students. But I believe that the essential step is to build up the comprehensive mental picture by testing it against what you know about the world, by means of a variety of crucial "thought experiments," and/or by conducting experiments, explaining them more or less successfully on the basis of the model and thus successively improving your conception of the model's assumptions, how they fit together and how they should be used.

The purpose of this essay is to exemplify and explore the details of this process of inquiry, helping build up a progressively more powerful ray model of light by means of several "thought experiments." Thus I will present, first, two questions about images formed by a pinhole and by a lens. Diagrams and detailed explanations of how to apply the ray model to these situations are included, but I invite you, before reading my explanations, to pursue the questions yourself, make diagrams and write your own explanations.

After these two questions, I present a third question – an intriguing and somewhat more challenging extension of the earlier questions. You are invited to construct your own response to this third question, as a way to gain experience with the inquiry process as well as additional insight into the ray model. My own diagram and explanation are also included. This essay concludes with my reflections on scientific models, the inquiry process, and the challenge of helping learners use scientific models effectively.

Here are the three questions:

Question 1. Have you ever observed an image on a screen formed by a pinhole? If so, you very likely noticed that the image was upside down. Can you explain this with a diagram and a model for how light travels? You can use Fig. 1 as a starting point for your drawing. Explain the assumptions you are making about how light travels. Are there any other features of the image you would expect using your diagram? In particular, how would you expect the image to behave if the screen were moved further away from the pinhole? See below for an answer to and explanation of Question 1.

Question 2. Have you ever observed an image formed on a screen by a convex lens? If so, you very likely noticed that the image was, as in the case of the pinhole, upside down. The lens image, however, differs from the pinhole image in two important ways: it is brighter, and it is "in focus" only when the screen is at one position; if the screen is moved closer to or further from the lens, the image becomes steadily more blurred. Can you draw a diagram and explain this on the basis of the ray model? You can use Fig. 2 as a starting point for your drawing. The key new assumption you must make here is that the lens *bends* the light, that is, the light no longer travels in a straight line when it goes from air to glass, or vice versa. See below for an answer to and explanation of Question 2.

In the spirit of inquiry, before proceeding to Question 3, I invite you to work out your own responses to Questions 1 and 2, and, if you wish, to consult the explanations and diagrams. When you are ready, proceed to Question 3.

Question 3. Suppose you were to use *both* the pinhole and lens, placed next to one another, to form an image on a screen (as in Figure 3)? Would you predict that the resulting image would be upside down or right side up? You are invited to make a prediction and to justify your prediction with a drawing based on Figure 3. Naturally, you will want to draw from your thinking about Questions 1 and 2 preceding. Whether or not you make a prediction that agrees with what is actually observed, the process and logic you use to arrive at your prediction, and your experience in making diagrams and in applying the ray model, will be a key foundation for gaining deeper insight into the nature of images and how to explain them using the ray model. See following for the observed result and an explanation.

Explanation of Question 1 – How is an image formed by a pinhole? Figure 4 illustrates the formation of an image by a pinhole. A bundle of rays diverging from point A near the top of the glowing object pass through the pinhole and illuminate the point on the screen at **A′**. A bundle of rays diverging from point **B** pass through the pinhole and illuminate the point on the screen at B′.

Now, why is the image upside down? Note that the rays from the top (**A**) of the object must travel downward in order to go through the pinhole and reach the screen, while the rays from the bottom of the object (**B**) must travel upward. In other words, the rays from the top and bottom of the object cross at the pinhole; this is what guarantees that the image will be inverted. In fact, you can pick any two different points on the object; the two rays must cross at the pinhole, and the image will be an exact copy of the object, but upside down.

To understand better how the image is formed, think about what would happen if there were additional pinholes in the foil. Each pinhole would give rise to a different image, and each image would be slightly displaced from one another, depending upon the location of the pinhole. These images would overlap and tend to wash each other out. If we were to take away the foil, there would be a huge number of overlapping images, which would result in a uniformly illuminated screen. Thus we can think of the pinhole image in two ways, either as the result of the light that goes through the pinhole to the screen, or as the result of the light that is blocked by the aluminum foil, allowing the image from the light that finds its way through the pinhole to become visible! Another way to say it is that the foil and pinhole are eliminating all of the other images and permitting rays that will form only *one* of the images to reach the screen.

It is important to note that we are assuming something here: that each point of the glowing object emits light rays traveling in all directions and that the rays from a point carry information only about that particular point and not about the image as a whole. Thus we are thinking about the process by which an image is formed as a kind of interaction between a source (the glowing object) and a receiver (the screen and then our eyes); the light rays traveling in straight lines are the way this interaction is transmitted from source to receiver.

It is also interesting to think about what will happen as the screen is moved further from the pinhole. Using the diagram, we can imagine the screen moving to the right, with the rays continuing to spread out. Thus we expect that the image will appear at any distance from the pinhole and that the *size* of the image (the distance from **A′** to **B′**) will be directly related to the distance between the pinhole and screen. In addition, we can see that the bundle of rays from A and from B *also* continue to spread out after they pass through the pinhole, so that the size of the "points" **A′** and **B′** will get larger as the screen moves to the right, showing that the edges of the image should become less distinct or fuzzier, and the image itself should become dimmer as the screen moves further from the pinhole. You can indeed verify (with a candle in a dark room) that an image from a pinhole behaves like this.

Finally, what effect will the *size* of the pinhole have? Again, we can use the diagram; if the pinhole is larger, a larger bundle of rays will pass through, but they will also be spread out over a larger angle. Thus we would predict that a larger pinhole would form an image that will be brighter but fuzzier.

Explanation of Question 2 – How is an image formed by a lens? Figure 5 shows a ray diagram for an image formed by a lens. The rays originating at A diverge until they strike the lens where they are bent so as to come back together at **A′**, which is the "image" of point A. Similarly for the rays originating at H. We have shown what happens to the rays starting from **A** and **H**, but other points (such as **B**-**G**) on the object also generate a cone of rays. In fact, you should imagine an infinite number of points all over the object; each point originates a cone of rays which strikes the lens and is bent so as converge at a distinct image point on the screen.

One can see immediately that the image is inverted: the image of the top of the object (**A′**) is at the bottom of the screen, and the image of the bottom of the object (**H′**) is at the top. In this diagram, the rays are not traveling in straight lines, as they were in the case of the pinhole; they are now changing direction at the lens.

But, why is the image inverted? As one can see from the diagram, the rays from **A** and **H** are, as in the case of the pinhole, crossing each other at the lens: the rays from A start at the top of the diagram and end at the bottom; the rays from **H** start at the bottom and end at the top.

The upside-down image is a direct result of the rays from the top of the image crossing the rays from the bottom. This is because all of the rays that reach the screen must pass through a limited opening, or an aperture. This aperture is defined by the opaque material around the lens. Thus the inversion of both the pinhole and lens images is due to the existence of a limited aperture: in order to get through the aperture, all the rays from any point on the object must cross the rays from other points on the object as they go through the aperture. In the case of the lens, however, the cone of rays leaving each point of the object, which reaches the screen and forms the image point, is indeed much larger than in the case of the pinhole. In Figure 5, I have illustrated this by showing several small cones of rays leaving each point; each of these cones is equivalent to the cone that might reach the screen through a pinhole.

If the lens itself doesn't invert the image, what *does* it do? The lens changes the direction of the rays, so that all the rays originating from a particular point on the object come together, or converge, at a particular point (as at **A′** through **H′**). This is the essential feature of the lens – rather than the cone of rays from a particular point on the object continuing to diverge after they pass through the aperture, as with the pinhole, all the rays originating from a particular point of the image are bent as they pass through the lens so as to converge to a single point. As you would expect from the diagram, if a screen were placed at the point where the rays converge, the image formed is brighter than the pinhole image. In addition, the image formed by the lens will be sharper than the pinhole image. However, the image formed by the lens is only sharp when the screen is at a specific distance from the lens; in the diagram, you can see that if you were to move the screen to the right or left the image would very quickly get fuzzy and out of focus. In addition, Figure 5 shows that the brightness of the image is related to the size of the lens (the diameter of the aperture).

In summary, the pinhole prevents most of the light from reaching the screen; the rays that pass through the keyhole must cross all the other rays at the pinhole and as a result form an inverted image. The rays continue to diverge after passing through the pinhole; thus the image is relatively dim and fuzzy. The image can be displayed on a screen located at any distance from the pinhole; this is a property that is unique to the pinhole image.

On the other hand, the lens also forms an inverted image. This can be explained in the same way as the inversion of the pinhole image: all the rays must pass through the lens opening (the aperture), just as in the case of the pinhole. Even though the lens aperture is larger than the pinhole, the rays from any point on the object must still cross the rays from all other points on the object, and this fact alone guarantees an inverted image. The crossing of the rays, and the inversion of the image, thus, is the result of the existence of the aperture, not of the action of the lens.

The lens does not really invert the image; the function of the lens is to bend the light that strikes it so that the rays that diverge from each point converge at a point located on the other side of the lens. This, and not the inversion of the image, is the distinctive contribution of the lens. The lens bends each cone of light that originates at a point on the object so it converges onto a specific point on the screen; the pinhole doesn't do this, it selects out part of each diverging cone, but the cone continues to diverge after it goes through the pinhole. This converging, or focusing, action is what makes the image formed by the lens brighter and sharper than the pinhole image; however, the lens image will be sharply defined only if the screen is at a specific distance from the lens.

Explanation of Question 3 – Is the image formed by a pinhole and a lens together upside down or right side up? Now, how about the image formed by a pinhole in front of a lens? Many students reason that the image would be inverted twice and thus appear right side up on the screen. However, in fact, the image appears *upside down!* Before reading the explanation following, you are encouraged to revise your own explanation and diagram as necessary to account for an upside-down image.

Figure 6 illustrates the situation. The image is right side up because the lens and pinhole together act as a single aperture. Therefore, the rays cross just once and the image is still inverted.

Another way to think about this is that the pinhole simply restricts the aperture of the lens, so that fewer rays reach the lens. In a sense, the lens is now using the aperture of the pinhole rather than the wider aperture of the lens itself. Therefore, the image created by a pinhole in front of a lens should be the same as the image created by the lens itself, except not nearly as bright, since so many of the rays are prevented from reaching the lens. As you can see in the diagram, however, the rays passing through the pinhole will now, because of the lens, be brought together at a point, rather than continuing to diverge as they did without the lens. So the image should also be sharper than the image created by the pinhole, in fact, just as sharp as without the pinhole, but with fewer rays and thus dimmer.

A third way to think about it is to visualize the result if there were several pinholes in the aluminum foil. As explained earlier, if there were no lens, there would be several overlapping images, one from each pinhole on the screen. However, the lens works, as described, to bend all the light coming from a particular point on the object so that it converges at a particular point on the screen. As a result, when multiple pinholes and a lens are used, the images due to the various pinholes would not be distinct overlapping images tending to wash each other out; all of them would fall exactly on top of one another, thus forming a much brighter image.

There is at least one more useful insight we can glean from Figure 6: the image formed will have one other characteristic due to the pinhole: the image is not located only at one location. The screen in Figure 3 can be moved toward and away from the lens and the image will still be visible. The cone of rays is restricted by the pinhole, so that the image can be viewed at a much wider range of distances from the lens. This is known in photography as the relationship between "depth of field" and aperture. That is, if the camera aperture is small, then the focus is not too important, and objects at various distances from the photographer will still appear sharp in the photo. On the other hand, if the camera aperture must be opened wide (because of little light), then getting the focus right is very important, and there will be a relatively narrow range of distances from the camera within which objects will appear sharp in the photo.

Finally, I suggest a final question (to which I will not provide the explanation!):

Question 4: Can you find a way that a pinhole and a lens *can* be arranged so as to produce an image on a screen that is right side up?

Reflections on scientific models and on inquiry.

The fundamental properties of light that are usually assumed as part of the ray model, and which we have tried to build up on the basis of "thought experiments" are as follows:

> Light travels from a source to a receiver in straight lines, called "rays."
>
> A single ray does not carry the whole image, just "information" about the specific point it came from.

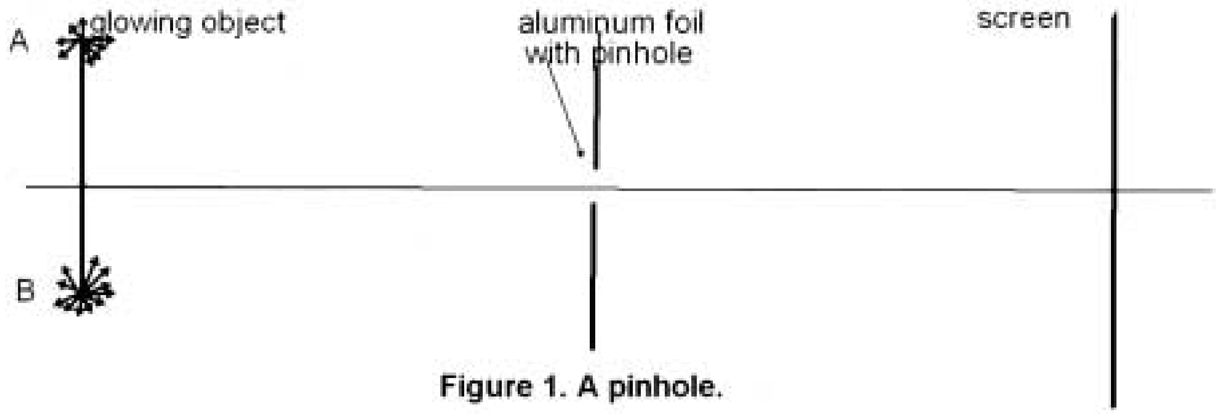

Figure 1. A pinhole.

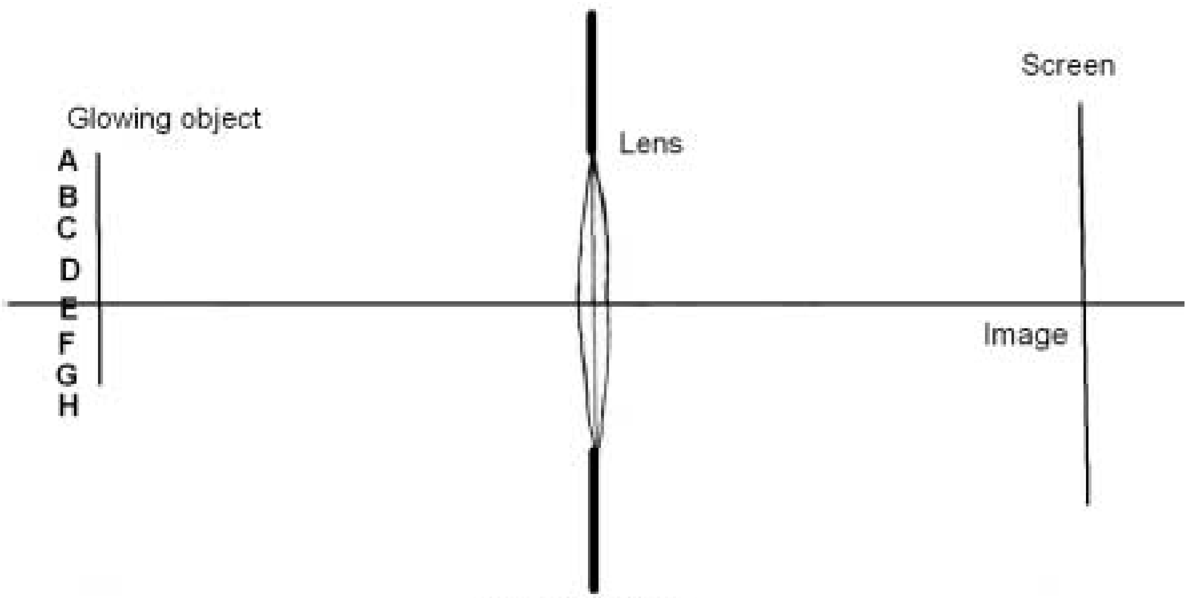

Figure 2. A lens.

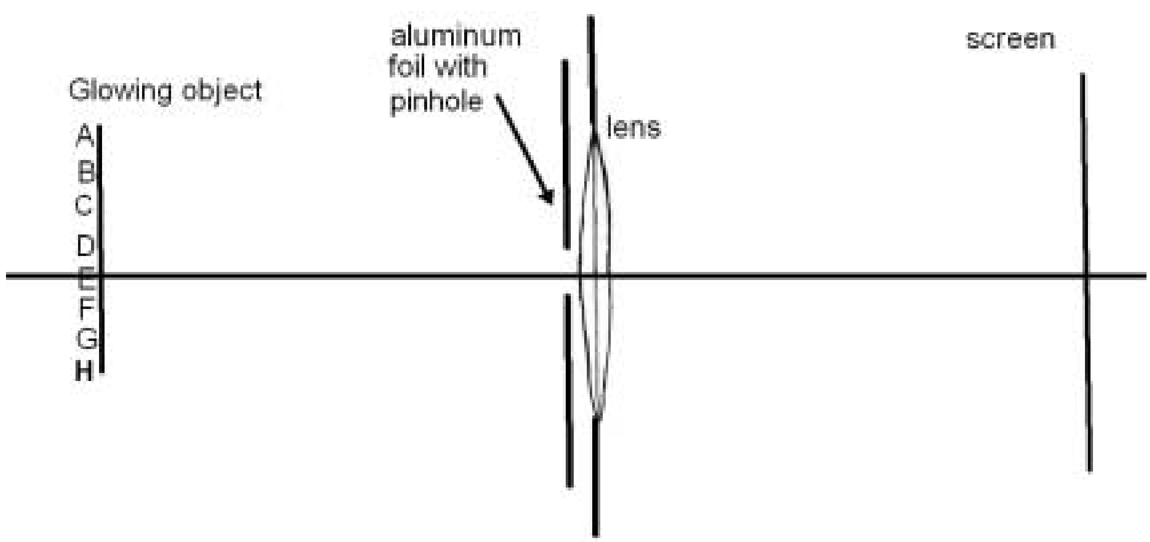

Figure 3. A pinhole and a lens.

Rays go in straight lines unless acted on by something else; when passing from one substance to another, light bends, or changes direction in a consistent way.

Each point on a source emits many rays in all directions

Rays can overlap or cross each other without affecting each other.

These assumptions, plus a few others, collectively form the ray model of light. Using this model effectively requires being able to actively participate, to some degree, in the process of abstracting these ideas from the actual phenomena, as well as keeping straight the differences among (in order of descending generality): the underlying assumptions of the model, the diagrams and the actual observed phenomena.

The process of inquiry in science involves carrying out your own experiments or recalling your own experiences with a particular phenomena, constructing your own mental model, using it to understand what you have observed and, most important, using it to predict the results under new conditions or in changed circumstances. This process is not a "one-way" process – rather, it involves encountering contradictions and changing your thinking in some way to resolve the problem. This change in thinking can involve modifications at any level of abstraction: in the underlying assumptions behind the model, in the reasoning or logic involved in applying the model, in the specific diagrams or equations used in applying the model, or even in the way the observations of the real world are carried out. Memorizing the assumptions of a particular model is of little help in building up one's own mental version of a scientific model and in using the model effectively. Rather one must actively engage with the physical world, subjecting the model to critical tests and modifying it as necessary to account for and explain one's own experiences and carrying out a variety of "thought experiments" based on such experiences.

The examples given involve using the ray model to account for the appearance of the pinhole and lens images. A deeper understanding of the model, and more careful analysis were required in order to resolve the paradoxical result when the pinhole and lens were used together, but the ray model itself did not fail or have to be modified. However, no scientific model, including the ray model, is universally applicable, and there are always situations in which a given scientific model will fail.

In particular, the ray model fails to predict phenomena that are observed when the sizes of the pinholes and other apparatus are very small, less than about one ten thousandth of an inch. In such cases, another model, the wave model, can be used more successfully. To account adequately for other situations, we have had to modify the ray model and to think of light as consisting of individual particles that are transmitted from source to receiver.

The generally accepted current theory of light, known as quantum electrodynamics (QED), was developed in the 1940s by Feynmann, Tomonaga and Schwinger. Quantum electrondynamics does not rest on a simple intuitively satisfying mental image and a direct analogy with everyday phenomena in the same way that the ray and wave models do. On the other hand, QED has a set of clearly stated, logically consistent assumptions and resulting mathematical equations that can be used in an agreed-upon way to yield accurate predictions about the behavior of light. But it is not correct to say that QED consists only of abstract mathematics. Just as with the wave and ray models, there are a variety of practical images and "heuristics," shorthand ways of thinking about QED, analyzing phenomena, generating an intuitive feel of what is likely to be the result and guiding us toward generating the appropriate equations. For example the "Feynman diagrams" offer much the same power and help in thinking about how to relate QED to what is observed as our diagrams provided for working out how to apply the ray model to pinholes and lenses.

Thus we see the continuing importance of simplified mental "images" and constructs as the key foundation for scientific modeling and for the more formal assumptions and equations of scientific theory. This also brings out the critical role of inquiry (trying out various assumptions, doing experiments, seeing how well you can explain what you find or what you have previously observed) in developing effective mental constructs and strengthening the ability to use scientific models effectively.

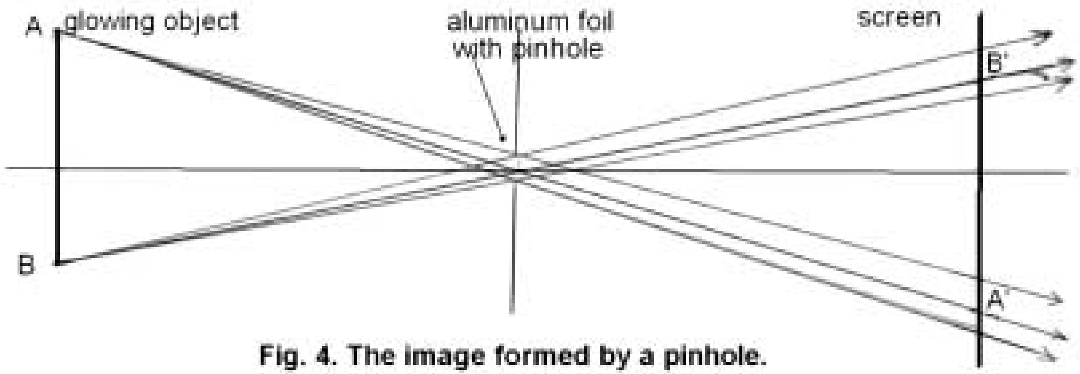

Fig. 4. The image formed by a pinhole.

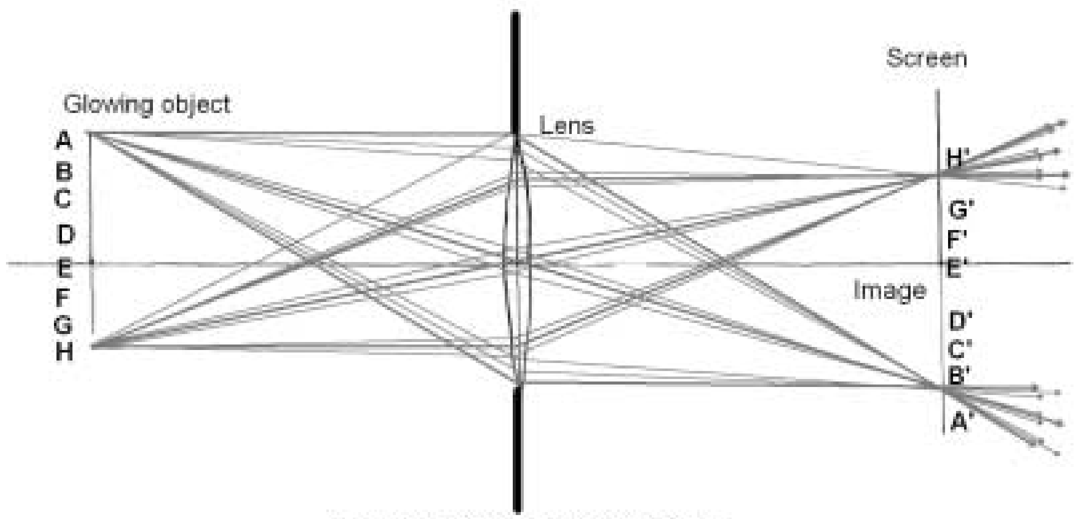

Figure 5. The image from a lens.

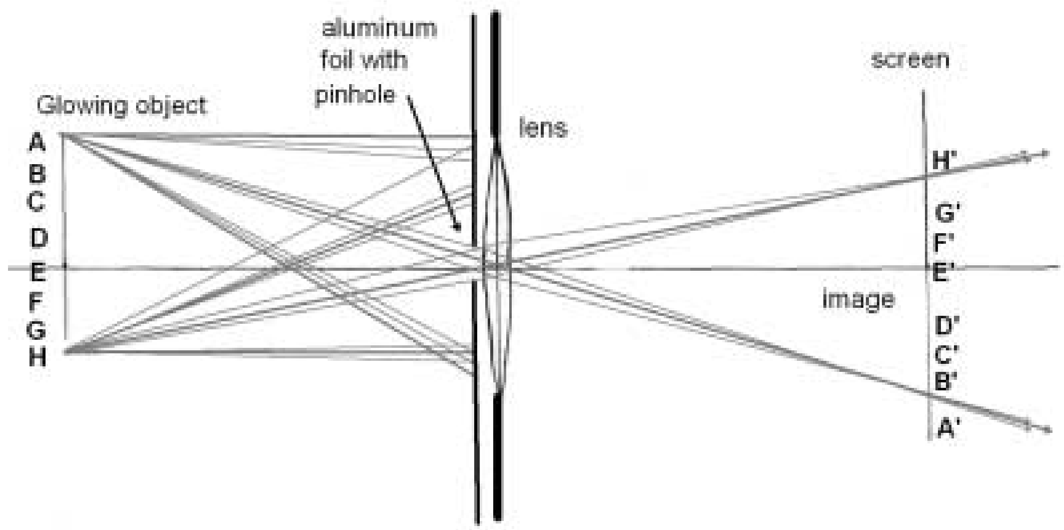

Figure 6. The image formed by a pinhole and a lens.