With the demise of traditional empiricism, the project of reconciling scientific realism with an acceptable scientific empiricism becomes pressing. This paper will address two questions relevant to that project. The first is: Is it possible to make a reliable and justified inference from the data to the properties that generate the data? The second is: Can the use of techniques to correct ‘raw data’ be an epistemologically useful practice? The answer to both of these questions is ‘Yes’. The first affirmative answer supports a form of scientific realism. The second affirmation suggests that a simple foundationalist attitude towards data is untenable. I shall discuss these questions within the context of a certain kind of causal-computational instrument, X-ray computed tomography scanners. These can provide us with a better insight into the answers to our questions than can purely causal instruments, such as optical telescopes, because the use of models and inferences within them is explicit, reasonably well understood, and in some cases can be assessed on a priori grounds. An additional benefit is that we can see how much more complex is scientific empiricism than traditional empiricism while allowing the former to act as a justifiable basis for scientific knowledge.

By ‘scientific realism’ I mean a selective position that asserts the existence of specified entities for which there is substantial scientific evidence but denies that all entities referred to in successful scientific theories exist. A realist position in which all non-mathematical terms that occur in successful scientific theories, models, or simulations refer to entities in the natural world is indefensible. Idealizations, quantities introduced for computational convenience, stylized facts, the subject matter of toy models, and many other things occur in scientific representations, often with an explicit denial by their users that they are committed to the existence of anything that corresponds to such conveniences. Perfectly smooth continuous surfaces on macroscopic objects, artificial viscosity, that the complementary cumulative probability distribution of real returns in financial markets satisfies a power law, and artificial insects in a model of social cooperation are,
respectively, examples of the four types of unreal components listed above. Regarding empiricism, an acceptable scientific empiricism is one that accords a special place to data with empirical content that is gathered and processed according to approved scientific practice but that does not uniformly give priority to data that are accessible to human sense perception. To so restrict oneself is to deny that instruments can in some cases provide epistemically superior data.

If one is a realist, at least some data have empirical content that is generated by something distinct from the data themselves and moving from the data to the generating source requires an inference. It is these inferences at which traditional empiricists balk. Logical constructions from data reports are acceptable, inferences to unobservables are not. From the realist perspective, in conditions where well established knowledge of the processes that generated the data is available, that attitude is unnecessarily risk-averse.

An epistemological principle, which we can call the Data Principle, is frequently used in traditional empiricism: The closer we stay to the raw data, the more likely those data are to be reliable sources of evidence. Here is a typical presentation of that view: ‘[…] we must observe [the world] neutrally and dispassionately, and any attempt on our part to mould or interfere with the process of receiving this information can only lead to distortion and arbitrary imagining’ (Lacey 1995, 226). Although the Data Principle had sound reasons for adoption by traditional empiricists on the grounds that any theoretical interpretation of data undermined their objectivity, and that any inference from raw data reduced the certainty associated with that data, an empiricism suitable for science should reject the principle. It is not just, as Suppes (1962) argued, that the judicious use of models is necessary to connect theory with data, although that is certainly true. It is that raw data coming from certain kinds of scientific instruments must be corrected and understood to improve the representational accuracy of these instruments. The better we are able to correct the raw data, the better the representations from those instruments will be.

1 X-ray Computed Tomography

I take as my running example X-ray computed tomography (CT) imaging devices. Although other imaging modalities such as positron emission tomography, magnetic resonance imaging, and single positron emission tomography use related computational techniques and many of the conclusions reached here generalize to those instruments, the physics underlying those devices is different from that occurring in CT scanners and this requires subtle but important modifications to the methods here described.
The mathematical methods behind CT image construction are complex and probably unfamiliar to many philosophers. For one of the few philosophical discussions see Israel-Jost (2011). A good introduction to the scientific and mathematical aspects of CT is Kak & Slaney (1988). I shall provide what I hope is a helpful outline of the main techniques but I stress that the current generation of scanners contains many more construction and correction algorithms than I can describe here. It is also worth keeping in mind that similar mathematical techniques are applicable in other sciences such as astronomy and geophysics in some of which, unlike medical imaging, a direct check on the accuracy of the images is not possible. Thus, although this direct accessibility to the target objects is epistemologically helpful, it is not essential in most cases to drawing the conclusions reached here.

One aspect of CT scanners that is relevant to selective scientific realism concerns the physical processes used by these instruments. X-rays were first systematically studied in 1895 and in the many years since, we have learned an enormous amount about their properties. To suggest that we do not know that they exist or that we should remain agnostic about their existence is to allow evidential concerns about genuinely problematical cases, such as the possible existence of dark energy, to affect our attitude towards entities that are of the same kind as well-understood entities such as visible light. Perhaps some of what we claim to know about X-rays is incorrect but it is sufficiently unlikely that most of what we claim to know about X-rays will turn out to be false that to reject realism in this realm is to engage in a level of scepticism that is unwarranted in scientific pursuits. We can accommodate those worries to a certain extent by adopting an ontology of properties rather than objects. That is, we take X-rays to consist in a cluster of property instances. Then claims concerning what X-rays ‘really are’ can be fallible in the sense that a few of those properties may turn out to be absent when X-rays are generated while acknowledging that there exist well established properties such as diffractibility, energies in the region 120 eV to 120 keV, being associated with an increased risk of cancer, and so on.¹ That is the realist side of our position with respect to features of the instrument that are present regardless of which target the instrument is applied. The selective side of the realism enters when we consider the image of a specific target.

The reason for having to be selective about our scientific realism is the presence of artifacts of the instrument; features of the output image that are

¹The exact classification of X-rays is in any case a fluid matter. The older criterion that distinguished X-rays from γ rays on the basis of wavelengths has largely been supplemented by the criterion that γ rays originate in the atomic nucleus whereas X-rays have an electronic origin.
the result of the physical or computational processes that are used in image generation and that do not represent features of the target. Not everything displayed in the outputs of CT scanners corresponds to something real in this sense: the representational content of at least one part of the image does not correspond to the real structure of the associated part of the object being imaged even if all parts of the image have their causal origins in the target.

Data from medical imaging devices are, as with other types of instruments, about specific individuals. Although in some cases the primary purpose is to make an inference about the individual being imaged, in other cases, especially in academic settings, inductive inferences to a wider class of subjects are the goal. Yet even in cases where a specific individual is the object of study and one might have worries that this specificity rules out the use as scientific, the concerns are misplaced. Galileo’s telescopic observations of the Moon were no less scientific than were his observations of bodies moving on inclined planes. The broader point is that theories may need to be general but scientific data do not.

The basic set-up for X-ray computed tomography consists of:

(a) a target object, located within a fixed coordinate frame that I call the target frame. Target frame positions are represented by Cartesian coordinates ($x, y$). In practical applications the target object will usually be a biological entity with an internal structure that differentially attenuates X-rays that pass through it. For example, bone attenuates the X-rays to a greater degree than does muscle. Physical calibration of the instrument is implemented using specially designed targets with known attenuation coefficients such as water and computational calibration is carried out on phantoms — artificial figures — such as the Shepp-Logan figure.

(b) M sources of X-rays and an array of M detectors. The sources, regularly spaced $\Delta r$ apart, are attached to a frame that rotates around the target within a two dimensional plane. Attached to this frame is a row of detectors also spaced $\Delta r$ apart, each of which is directly opposite one of the sources. This rotating frame defines the detector frame and is represented by the polar coordinates ($r, \theta$). To keep things as simple as

---

2Calibration using phantoms has a similar motivation to those at work in random number generators — the properties of mathematically generated random numbers are known exactly, while those generated from physical processes are not. Analytic results are available for projections of the ellipses used in the Shepp-Logan phantom.
possible, I consider the situation in which the beams of X-rays emitted by the sources are parallel.\footnote{Fan beams require greater mathematical sophistication but do not introduce any essentially new philosophical issues. Continuous 3-dimensional tomography uses a spiral path for the detector frame for which there are important additional complications that will not be considered here. Even if the specific instruments described here are no longer used in practice the conclusions drawn about the compatibility of selective scientific realism and scientific empiricism remain.}

After each firing of the X-ray sources, the detector frame is rotated around the target by $\theta$ degrees. There will be $N = \pi / \theta$ values of $\theta$ between 0 and $\pi$. Values between $\pi$ and $2\pi$ in the planar case are unnecessary because of symmetry. An important part of the computations involved in CT concerns transformations between the target and the detector frames.

2 Empirical content of data

Let $f(x, y)$ represent the value of some spatially distributed quantity. The central task is to calculate the values of $f(x, y)$ at various points within the target frame using values of detector counts. Before proceeding we need to disambiguate the use of ‘empirical content’. We can attribute empirical content to data considered as concrete entities that occur as elements of causal processes and also to data considered as representational entities. The latter can take a number of different forms including perceptual entities, linguistic entities, graphical images, and so on. (See Humphreys (2000); Kulvicki (2010); Vorms (2012) for discussions of different representational formats.) In what follows context will determine when notation picks out a concrete datum and when it denotes a representational datum. Next, define a datum as either a) any entity of the form $R(a_1, a_2, ..., a_n, s)$, where $R$ is a (possibly compound) concrete relation or property and $s$ is the spatiotemporal location of the property instance or b) a representational correlate $R(a_1, a_2, ..., a_n, s)$ such as a predicate instance or a set of spatial relations between pixels forming an image. A common use of ‘empirical content’ asserts that a concrete datum has empirical content just in case it can be accessed by whatever means the empiricist tradition in question considers permissible. I leave these means deliberately unspecific in order to cover different versions of traditional empiricism including sense-data empiricism, sub-species of logical empiricism, and constructive empiricism. More liberal is the position taken here that a concrete datum does not have to be accessible by the unaided human senses but it must be detectable by some reliable scientific instrument that is currently available. The empirical content of a datum can then be identified with the property instance, a position that is argued in detail in Humphreys (2004, chap. 2). This definition has the
consequence that data can lose their empirical content. One example would be an electron microscope image stored in an obsolete medium for which readers no longer exist. Although one could take the view that the data retain the disposition to be accessed, this would require abandoning the position that what is important for science is what can be observationally accessed in practice.

These degrees of strictness cover a division within empiricist traditions. One branch is opposed to rationalist methods of inquiry and so the approved means of access explicitly preclude the exclusive use of non-perceptual intuitions, a priori reasoning, and similar methods to arrive at knowledge. The other branch of empiricism is opposed to realism; the usual foil is scientific realism so that the use of scientific theories and instruments that, respectively, refer to or detect unobservables, are impermissible under traditional empiricism. All varieties of empiricism should fall into the first branch if the requirement that our knowledge has empirical content is correctly formulated but our goal is to avoid the second branch. A priori methods do play a role in empiricism in that the domain of empirical content is closed under computable transformations, where ‘computable’ means ‘computable by an existing device within time constraints set by the nature of the problem’.

Data sets can be adjusted, a process that includes throwing away data and inserting data. The topic of whether to keep or throw away anomalous data is an important one but it has been discussed elsewhere and I shall not address it here. Data can be added to a data set either by providing additional raw data, by generating artificial data, or by transforming raw data. Transforming data involves taking existing data and performing physical or formal transformations on them. In the case of CT instruments, this can involve coordinate transformations, interpolations between existing data values, the use of other mathematical operations on data, and applying physical operations to data.

Physical and formal transformations do not always preserve reference to reality even when they preserve empirical content. Consider the representation of the scalar value of the velocity of a macroscopic object, \( v \). This is a datum with empirical content, we know how to measure it, and it represents something real. Now consider the variable \( v^{17.34} \). We know how to measure what it represents – measure \( v \) and apply the appropriate mathematical transformation. So it also has empirical content. But the variable \( v^{17.34} \) does not correspond to anything real according to the current state of scientific knowledge. A specific value of \( v^{17.34} \) can, by accident, correspond to something real, such as the exact value on another occasion of \( v \) but that does not entail that there is a property corresponding to \( v^{17.34} \). Compare this with length \( l \). If we take a particular value for the length, say 2 feet,
and square the numerical value, we have another value of length, 4. If we take the numerical variable \( l^2 \), that will represent area, which is clearly a different property from length. Although empirical content is preserved by mathematical transformations, they do not always preserve the property of being observable under traditional criteria. If \( l \) represents the length of an object that is in the domain of the observable, then \( l^{-100} \) will not be an observable. Many empiricist traditions have held that any representation that is explicitly definable using only observable terms will itself count as an observable term, perhaps because it would be eliminable via the definition, but that reasoning fails prey to the example just given.

Now consider a physical transformation of the datum \( v \) by means of an interaction with some physical system. Under some such transformations, the empirical content will be lost altogether in the sense that the interaction will convert the datum into a property instance that cannot be detected by current technology. In other cases the empirical content will change as the interaction changes one physical property into another. This is why identifying the empirical content of a datum with its source is not to be recommended, since data from the same source can be physically transformed in different ways.

Which functions of data are taken to represent instances of real properties cannot be read off the representations. The selection is made by the users of the representation not, as social constructivists would have it, on the basis of a social consensus, but on the basis of evidence that is often subject matter specific.

The overall moral is then that when asking whether a datum subject to transformations has empirical content, we must pay careful attention both to the type of transformation (physical or mathematical) and to the type of data (concrete or representational).

### 3 Basic CT computations

Returning to the general case of CT procedures, \( f(x, y) \) represents the attenuation coefficients for X-rays at points \((x,y)\) in a two dimensional plane through the three dimensional object. \( f \) allows us to calculate the amount by which a given X-ray beam is attenuated over a distance \( \Delta y \) where \( \Delta y \) is the dimension of one pixel in a discretization of the target space. Although the degrees of attenuation are themselves a function of the densities of various types of tissue and these densities in turn correspond to the presence or absence of various anatomical features, for simplicity I shall deal only with the function \( f \) and its values. The underlying point is, nevertheless, important; what is ‘observed’ is a function of the spatial distribution of attenuation coefficients of the target and an interpretation or inference is
needed to move from this spatial distribution of coefficients to the existence of biological features. Standard visual images consist in a spatial distribution of reflectances in the optical range; here I am generalizing this concept to allow an image to consist in the spatial distribution of any physical quantity. I retain the spatial aspect of images in part to maintain a connection with the idea that these are imaging devices producing a visual output but the values of the function $f$ could be displayed as a matrix of numerical values or in some other representation.

Call a raw datum a datum to which no transformations external to the instrument have been applied – that is, the datum is part of the instrument’s direct output. What is counted as a raw datum thus depends upon where the boundary of the instrument is drawn. In the present case, a raw datum consists of an X-ray count at one detector. For a given value of $\theta$ the elements of the raw data set are the counts received at each of the M detectors at positions $r_1$, $r_2$, ..., $r_n$. From these and knowledge of the initial intensities of the X-rays at the source it is easy to calculate the total attenuation of the X-ray beam along each of the M parallel rays. The central task is to reconstruct the values of $f(x, y)$ at specific points along those rays. The most common method of reconstruction uses filtered backprojection.

In broad outline, filtered backprojection contains these steps:

**Step 1** Calculate the total attenuation along a given ray between a source and its detector by integrating the values of $f$ along that ray

**Step 2** Convolve these spatial projections with a filter (a weighting function) to compensate for geometrical frame changes

**Step 3** Fourier transform these convolutions into the frequency domain to facilitate computations

**Step 4** Compute the convolutions in the frequency domain

**Step 5** Inverse Fourier transform the results back to the spatial domain

**Step 6** Compute the inverse Radon transforms in the spatial domain to arrive at values of $f(x, y)$ at the desired points $(x, y)$ within the target frame.

We see here that the ‘observed image’—that of a human hip, for example—is an entity that is constructed from classically unobservable data using explicitly considered transformations. Whereas sceptical doubts about unobservables are appropriate in cases where these transformations are unknown, in the current case a refusal to distinguish between real but traditionally
unobservable entities and processes and non-existent and traditionally un-
observable entities and processes to construct corrections to the image is
epistemically counter-productive.

Before discussing the details of these steps, some general issues concerning
the switch from continuous to discrete models is necessary. In the idealized
case where we have a continuous set of parallel beams and a continuous set
of detectors, we would have the entire profile of projections along \( r \). In the
actual case, we have to deal with these constraints:

1. The measurements consists in a discrete set of \( M \) equally spaced point
   values along \( r \) and this gives rise to the need for discrete models, with
   the accompanying need for approximations.

2. The discrete data set gives rise to an underdetermination problem
   that must be addressed.

3. Realism requires us to identify artifacts which can arise either from
   physical or from computational sources. To avoid physically gener-
   ated artifacts, correction models to supplement the basic model are
   required because of noise and systematic bias in the measurement.
   Correction models to compensate for the kind of computational er-
   rors discussed below are also required.

This transition between the idealized methods that use continuous math-
ematics and the discrete models used to represent the physical situation
requires attention. In the present case, the spatially continuous function \( f \)
is rendered discrete both by the use of pixels and by a truncation method
such as a finite series approximation to the values of \( f \). These finite approx-
imations are important because they can often make an inverse inference
problem ill-posed. An ill-posed problem, in cases where a unique solution
exists, is one in which small changes in the data can result in large changes
in the solution values. In such cases the discrete approximations must be ex-
amined in order to avoid an originally well-posed problem becoming ill-posed
so that small errors in the data lead to large errors in the solution. This
mathematical pitfall is a major difference between causal-computational and
purely causal instruments.

We can now consider each of the six steps outlined above.

**Step 1** Consider the detector frame when it is oriented at an angle \( \theta \) to
the target frame. Each ray can be represented mathematically by the line
parameterized by \( r \) and \( \theta \): \( L_\theta(r) = \{(x, y) : r = x \cos(\theta) + y \sin(\theta)\} \) where
\( r \) is the radial coordinate. The total attenuation along the line \( L \) is given
by \( \int_{L} f(x, y) dL \). This represents projected values of \( f(x, y) \) along the ray
orthogonal to the \( r \) axis of the detector frame when it is oriented at angle \( \theta \)
to the target frame. As a reminder, the empirical content of the raw data are the X-ray counts; from those one can calculate the total attenuation value along a given ray. To calculate this value using the target frame coordinates we have:

\[
(1) \quad P_\theta(r) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} f(x, y) \delta(x \cos(\theta) + y \sin(\theta) - r) \, dx \, dy
\]

which is the Radon transform of \( f \) over \( L \). \( P_\theta(r) \) is the X-ray intensity at the detector located at \( r \) in the detector frame. The coordinate transformations used on the right hand side of (1) do not alter the empirical content of the raw data; they are simply different representations of the same data set.

**Step 2** Given \( P_\theta(r) \), it would seem that if we performed an inverse Radon transformation for every value of \( \theta \), we could reconstruct an image of the object. There is one important caveat to note here. Because all of the \( N \) values of \( \theta \) are involved in the inverse transform, a reconstructed value of \( f \) is a function of the data for all \( N \) values of \( P_\theta(r) \), that is of the entire data set. There is thus a holistic aspect to the reconstructed data and this has the consequence that defects in the data at one point in the target can be transferred to many other points in the reconstructed image. Once we move to the discrete versions of the backprojection method that must be used in practice, simple backprojections produce a star effect around point sources due to the finite number of projections and blurring of the image occurs. A generalization of this problem results from the fact that the backprojection method regularly spaces values along the radial lines centered on the origin within the detector frame, and within that frame the data become sparser as we move towards the periphery of the object. Because of this the final image, which is represented within the Cartesian target frame, becomes less well defined, a feature known as \( 1/r \) blurring, which affects all regions of the image.

To correct for these problems, before we back-project the projection must be convolved with a weighting function (called a ‘filter’) to give \( P_\theta(r) * h(r) \) This operation constitutes the filtered part of the filtered backprojection method.

The convolution \( f * h \) of a function \( f \) with a shift-invariant operation \( h \) is defined by

\[
z(t) = \int_{-\infty}^{\infty} f(t') h(t - t') \, dt'
\]

In contrast to the coordinate transformations, the convolutions do not result in the same local empirical content. Rather, they redistribute the empirical content over the detector frame in a way that corrects for the
different geometry of that frame. Globally, but not locally, the empirical content is the same and the convolution corrects for distortions introduced by the data transformations. We have now seen the extent to which data must be corrected in these instruments and in each case the corrections are epistemically advantageous rather than harmful.\textsuperscript{4}

Steps 3 and 4 We now desire to infer from the values of $P_\theta(r)$ the values of $f(x,y)$ along a plane lying within the object. The key to doing this is the Fourier Slice Theorem which asserts:

The Fourier transform of a projection of a function $f(x,y)$, taken at an angle $\theta$ to the target frame, gives a slice of the two dimensional transform $F(u,v)$ that subtends an angle $\theta$ with the $u$ axis in the frequency domain.

That is, if we take the calculated projections $P_\theta(x)$ in (1) above onto the spatial detector plane and Fourier transform them into the frequency domain, this gives the Fourier transform within the frequency domain along a plane passing through the target object as long as that plane also passes through the origin of the spatial X-Y coordinates. Thus, given the data values at the detectors for a given value of $\theta$, by performing a Fourier transform on those values, we can obtain the values of $f$ along a plane in the object.

This has the additional advantage that because convolutions are computationally intensive but a convolution in the spatial domain Fourier transforms to a multiplication in the frequency domain, it is a significant advantage to carry out these computations in the Fourier domain. At this point we are far from the raw data.

Steps 5 and 6 These steps require less comment than the previous steps but in the next section I shall discuss some of the modeling steps that are required.

\textsuperscript{4} $h$ is often the ramp filter which weights each value of the Fourier transform by $v$, the frequency in the Fourier domain. But there is a trade-off involved because this weighting increases the high frequency noise and to increase the signal to noise ratio the ramp filter is itself multiplied by a window function such as the Shepp-Logan, the Hamming, or the mode-p Butterworth. The backprojection is carried out in the spatial domain whereas the filter is added in the frequency domain and has different formal properties in each. For example, the Ram-Lak filter in the spatial domain becomes the ramp filter in the frequency domain. In the spatial domain the Ram-Lak filter is given by $h(r) = 1/2\pi^2[sinc(r/\tau)] - 1/4\pi^2[sinc^2(r/2\tau)]$. The Hamming filter does better for improving the signal to noise ratio but worse on reducing the star effect.
4 Inverse inferences and underdetermination

The process of inferring the values of $f$ at specific points from the total attenuation values is what is called an inverse inference problem. Underdetermination problems are inescapable for empiricist positions and inverse inferences are no exception. We can carry out Radon transforms only up to a constant of integration, but there is a much more dramatic underdetermination problem that illustrates how the use of discrete data sets requires choices in practice that are not required in the continuous case. The result is this (Herman 1980, 283–286):

Let $\text{Rad}([f(r, \theta)])$ represent the Radon transform of the function $f$ at $(r, \theta)$. Let $M, K$ be positive integers and for $1 \leq k \leq K$, let $(r_k, \theta_k)$ be distinct points in the interior of the target. Let $l_k$ be arbitrary real numbers. Then there is a continuous function $g$ such that for $1 \leq k \leq K$, $g(r_k, \theta_k) = l_k$ and for $0 \leq m \leq M - 1$ and for all $l$, $\text{Rad}[g(l, m\pi/M)] = 0$.

Thus given any function $f$ that truly represents the target property, there exists a function $f + g$ such that $f$ and $f + g$ have the same line integrals along each of the $M$ X-ray beams but $f$ and $f + g$ differ by an arbitrarily large amount at each of an arbitrarily large finite number of points in the image plane. The proof of this result rests on the availability of an oscillatory function such that its Radon transform is invariant under scaling and translation and at each of the $M$ line integrals has value 0. Philosophers will recognize this result as a proof in this specific application of the abstract point that any function is underdetermined by a finite data set. In practice this underdetermination does not prevent construction of a near-correct image. The reason is that by setting the number of detectors $M$ at a number beyond the minimum required by the Nyquist sampling theorem, it is unlikely that the target object will exhibit periodicities that replicate those that underlie $g$. In the case of a one-dimensional function, the Nyquist sampling theorem requires that in order to avoid aliasing – the appearance of artifacts in the image – values of the function must be sampled at least twice during each cycle of the highest frequency contained in the spectrum of the continuous function (see Buzug 2008, 135).

5 Artifacts and noise

An artifact is a systematic discrepancy between the real attenuation values and the values inferred from the measurements taken at the CT detectors. (For a detailed assessment of artifacts, see Humphreys 2013). Artifacts have a number of different sources, both physical and computational, and cor-

---

5Inverse inferences are a rich source of inductive knowledge, largely ignored by philosophers of science. I shall discuss the inductive aspects of inverse inferences in a future paper.
revisions need to be made to the data to eliminate or reduce these artifacts. The need to correct for artifacts does not distinguish causal instruments from causal-computational instruments because lenses are physically corrected for chromatic aberration by using compound lenses. There are many possible sources of error that require correction but here I shall discuss one representative case, beam hardening. This occurs when the mean energy of the X-rays increases as they pass through matter because the lower energy X-rays are absorbed at a higher rate. This means that the effective linear attenuation coefficient of tissue decreases with distance from the X-ray source because the attenuation coefficient is a function of X-ray energy. The overall effect of higher energy X-rays is an increased incidence of Compton scattering and an accompanying loss of contrast in the image. Correction methods for this are applied before the image reconstruction takes place.

We can use this example of beam hardening to illustrate how our access to real, rather than artifactual, features changes with time. Algebraic and statistical reconstruction methods are superior to filtered backprojection in treating some cases of beam hardening but were not used for many years because of the excessive computational loads they require. With advances in technology, they are becoming feasible in practice and with them an increased ability to distinguish real features from artifacts. Data based knowledge is not founded upon a static epistemic basis as is traditional empiricism.

In addition to artifacts, noise is a factor in almost all instruments of this type. At a minimum there is a tradeoff between the achieved sharpness of the constructed image and keeping the signal to noise ratio at an acceptable level. For example, the correction factors used to reduce 1/r blurring introduce increased noise because the filter increases high frequency components in the Fourier domain. In optical instruments there has always been a tradeoff between different types of optical aberration and it is impossible to simultaneously optimize these errors for all wavelengths of light or over the entire surface of most lenses for monochromatic light. Here we have a similar situation but for computational corrections. The fact that raw data are not ideal should not prevent us from correcting them.

Suppose that the true values of a quantity y are given by a probability distribution $f(y)$ and that the noise distribution is given by $g(z)$. If we assume that the noise is stochastically independent of the value of $f$, then the probability that a given value of $f$ will be shifted an amount $z - x$ to the value $z$ by the noise contribution $g$ is $f(x)dx, g(z - x)dz$. Then, integrating over all the values of $x$ that give rise to $f(x)$ gives $\int_{-\infty}^{\infty} f(x)g(z - x)dx$.
= f * g = h(z).\(^6\) \(h(z)\) is the value of the output of the instrument and will ordinarily have a greater spread of values than \(f\) and be biased if \(g\) is biased. If \(f\) is a delta function, so that there is no stochasticity in the underlying physical variable, then the convolution places a copy of the noise distribution at each value of that variable.

Whereas the usual representation of error in linear models is with an additive function, the convolution is a general representation of how system values and noise combine. More important for us is the inverse of convolution, deconvolution. If we have the output distribution \(h\) and have a good model of the noise function \(g\), then in certain cases we can recover the system distribution \(f\). The important point is that in some cases we have knowledge of the distribution of noise in particular parts of the instrument and since we have access to the output distribution, we can compute the system distribution. Many of the philosophical discussions of noise operate under the assumption that all we have available is the pattern of data that constitutes the output from an instrument. This assumption is false in some cases. Because we have constructed these instruments, we are in a different epistemic relation with respect to noise originating with them compared to noise originating from a naturally occurring system such as a galaxy. Instruments are not experiments but the two share the feature that strict controls can be placed on independent variables and on some sources of noise.

This is one more reason why knowledge of how the imaging device works is crucially important in improving its accuracy. In the early seventeenth century, Leewenhoek and Galileo could construct optical microscopes and telescopes, respectively, using trial and error to produce a reasonable image while having nothing close to a correct theory of how they worked. In contrast, it would be impossible to construct a CT scanner without a considerable amount of knowledge of applied mathematics.\(^7\)

---

\(^6\)The convolution does not add the noise to the true value, nor is the convolution the composition of two point functions. The notation here is also important: \(z\) and \(x\) are representations of the same physical variable but will in many cases have different values for that variable.

\(^7\)Thanks are due to audiences at the Knowing and Understanding Through Computer Simulations conference, Paris; the 14th Congress on Logic, Methodology, and Philosophy of Science, Nancy; the Plurality of Numerical Methods in Computer Simulations and their Philosophical Analysis conference, Paris; the Computer Simulations and the Changing Face of Scientific Experimentation, Stuttgart; and the Models and Simulations conference, Helsinki for helpful comments on talks related to this topic. I am especially indebted to George Gillies of the Department of Mechanical and Aerospace Engineering at UVA for correcting errors in the penultimate draft.
Bibliography


Paul Humphreys
Corcoran Department of Philosophy
University of Virginia
Charlottesville VA 22904-4780
USA
ph2a@virginia.edu