How we load our data sets with theories and why we do so purposefully

Guillaume Rochefort-Maranda

August 16, 2016

Contents

1 Introduction

In this paper, I compare theory-laden perceptions with imputed data sets. The similarities between the two allow me to show how the phenomenon of theory-ladenness can manifest itself in statistical analyses. More importantly, elucidating

the differences between them will allow me to broaden the focus of the existing literature on theory-ladenness and to introduce some much-needed nuances.

One of the main reasons why theory-laden perceptions have been extensively discussed in philosophy of science is that they pose an apparent threat to scientific knowledge. As a result, their very existence has been challenged, or dismissed as merely one of the many potential sources of error inherent in every experimental framework that we can learn how to control.

In sharp contrast, the topic of statistical imputation has received no attention in philosophy of science. Yet, imputed data sets are very similar to theory-laden perceptions, and they are now an integral part of many scientific inferences. Unlike with the case of theory-laden perceptions, the existence of imputed data sets cannot be challenged or reduced to a manageable source of error. In fact, imputed data sets are created purposefully in order to improve the quality of our inferences. They do not undermine the possibility of scientific knowledge; on the contrary, they are epistemically desirable.

Accordingly, I propose to substantiate the following three theses:

1. Theory-laden observations need not be treated as an inevitable nuisance. They can be created purposefully for their epistemic benefits.

2. Some scientific inferences cannot be understood and assessed unless we understand how and why our observations are theory-laden.

3. The theories loaded onto our observations need not play a role in the justification of our scientific models.

This paper comprises three main sections. In the first section, I explain the thesis of the theory-ladenness of perception and highlight the fact that it has been

cited as an obstacle to scientific knowledge. This will set the stage for a fruitful discussion of imputed data sets.

In the second section, I introduce imputed data sets and furnish an example based on a simulated data set. I also show why statistical imputation is important. The results presented in that section can be reproduced using the R code provided on the following websites: `http://grmaranda.weebly.com/` or `https://ulaval.academia.edu/GuillaumeRochefortMaranda/Papers`. In the third and last section, I compare imputed data sets with theory-laden perceptions. This provides the philosophical crux of this article.

2   Theory-Laden Perceptions: The Controversy

The expression "theory-laden observation" refers to more than one phenomenon (Bogen 2014). For example, when we say that our observations are theory-laden, we can mean that our observation reports are couched in theoretical vocabulary (see Chalmers 2013; Churchland 1988; Kuhn 1962 for examples). We can also mean that our theoretical conjectures have directed our attention towards a certain observable phenomenon (see Couvalis 1997; Brewer and Lambert 2001 for examples). Here, however, I wish to focus exclusively on the idea that our theoretical beliefs can determine, at least in part, the content of our perceptual experiences. This is the thesis of the theory-ladenness of perception.[1]

To be more precise, the theory-ladenness of perception is a thesis in cognitive psychology. It holds that perception is the result of an unconscious inference which is determined by top-down theory information and bottom-up sensory infor-

---

[1] I elect to single out this thesis is because I intend to compare imputed data sets with theory-laden perceptions only. Other versions of the phenomenon of theory-ladenness, although intrinsically interesting, are not relevant for my purposes.

mation(Brewer and Lambert 2001; Gregory 1997). This theory can be traced back to Hermann von Helmholtz (1866):

> Following von Helmholtz's lead we may say that knowledge is necessary for vision because retinal images are inherently ambiguous (for example for size, shape and distance of objects), and because many properties that are vital for behaviour cannot be signalled by the eyes, such as hardness and weight, hot or cold, edible or poisonous. For von Helmholtz, ambiguities are usually resolved, and non-visual object properties inferred, from knowledge by unconscious inductive inference from what is signalled and from knowledge of the object world (Gregory 1997, p.1122).

It is a thesis that is usually presented as a conjecture meant to provide an explanation for various optical illusions: "Some phenomena of illusion provide evidence for the uses of knowledge for vision; this is revealed when it is not appropriate to the situation and so causes a systematic error, even though the physiology is working normally" (Gregory 1997, p.1122).

One such illusion that is often mentioned in the literature is the Ames room illusion (see Papineau 1979; Gillies 1993). It involves a non-rectangular room such that a properly situated observer can perceive two people of similar size as if they were of different heights. A possible explanation for this illusion is that our experience is determined not only by the visual inputs (bottom-up information) but also by our background beliefs (top-down information) about the usual shape of a room. In other words, the explanation for the illusion is that our perceptions are theory-laden.

Donald Gillies, for instance, clearly adopts this kind of explanation:

> It is easy to understand what is happening here. The brain has the

4

choice of interpreting the visual input according to either of the following theories: T1: The two people are approximately the same size, but one is much further away because the room is an odd shape. T2: The room is the usual rectangular shape, and the people are different sizes. Here, T1 is correct, and T2 incorrect. Yet anyone, on seeing the Ames room for the first time, will unconsciously opt for T2. This is perhaps not surprising. Experience has made us all very familiar with the fact that rooms are nearly always rectangular in shape, while people often vary in size. Thus, on the basis of experience, T2 is better confirmed than T1, although T1 is in fact correct. [...] The Ames room shows very clearly that ordinary everyday observation is theory-laden (Gillies 1993, pp.143-44).

Some have even claimed that the credibility of such an explanation is enhanced by the existence of neural pathways between the higher cognitive centres of the brain and the visual modules in which the early stages of perceptions occur.

The thesis of the theory-ladenness of perception has been presented as a feature of scientific practice which can produce scientific knowledge (see Hanson 1958; Feyerabend 1993; Kuhn 1962 for examples). But it has also been more recently discussed by philosophers of science because of the apparent threat that it poses to the possibility of scientific knowledge:[2]

If all observation is theory-laden, the objectivity of scientific research might be undermined, for it seems that we may well be unable to tell

---

[2]Other version of the thesis of the theory-ladenness of observation have been advanced in support of scientific realism (Maxwell 2009). In this paper, however, I only address the literature on the theory-ladenness of perception that focuses on the potential threat that it poses to scientific knowledge.

whether our perceptions accurately capture aspects of the world. (Couvalis 1997, pp.11).

The issue of the role of top-down, or reentrant, neural pathways that transmit signals from the higher cognitive centers to the perceptual modules is important to the philosophy of science, since it is the existence of these pathways that is used as one of the arguments for the cognitive penetrability, and thus for the theory-ladenness, of perception [...], thus clearing the way for relativistic theories of meaning and scientific theories (Raftopoulos 2001, pp.S187).

On the views of perception that Fodor is out to resist, (theoretical) conception is capable of penetrating perception thoroughly. Theoretical commitments infiltrate observation. Consequently, theory-neutral observation is impossible, and scientists must decide between competing theories on grounds that are pragmatic and holistic at best – grounds that Fodor finds insufficient for a satisfactory defense of scientific rationality. (McCauley and Henrich 2006, p.81).

The problem is that if we always justify theories with theories, then we appear to enter an infinite regress of justification or a vicious circle of justification. It is thus not surprising that many philosophers have challenged the existence of theory-laden perceptions (see Fodor 1984) or downplayed their importance as a credible obstacle to obtaining scientific knowledge.

Jerry Fodor, most notably, has pointed out that we can be victims of optical illusions, such as the Ames room illusion, even if we know how the illusions work (see Fodor 1984). Therefore, even if our beliefs were capable of influencing the content of our experiences, that influence is clearly very limited and it is thus unable to adequately sustain the argument against the possibility of scientific knowledge.

Moreover, it has been argued that our perceptual experiences play a very minor role in many experimental frameworks. Scientists often merely observe digits or needles on a scale (see Fodor 1991). The visual stimuli in such environments are not ambiguous, vague, or anomalous. Yet the most convincing cases of theory-laden perceptions involve just such stimuli. Indeed, they are necessary to produce optical illusions such as the Ames room illusion. Hence, even if we acknowledge the existence of theory-laden perceptions in some contexts, it is doubtful that they have much impact on many of our scientific inferences, and they consequently offer a poor rationale for questioning the possibility of scientific knowledge:

> However, note that in all of the above cases the stimuli were either ambiguous, degraded, or required a difficult perceptual judgement. In these cases the weak bottom-up information allowed the top-down influences to have a strong impact on perceptual experience. [...] If the information to be perceived is whether a needle on an instrument is registering a 10 on a clear 1-10 scale it is unlikely the theoretical beliefs of the scientist will be able to override the strong bottom-up perceptual information. Thus the top-down/bottom-up analysis allows one to have cases of theory-laden perception, but not necessarily lead down the slippery slope to relativism. (Brewer and Lambert 2001, p.S179).

In fact, some philosophers now believe that theory-laden perceptions could at most be a hindrance to obtaining scientific knowledge in some limited research contexts in which the visual stimuli are poor. Here is an example:

> That our experience is untrustworthy under poor conditions does not imply that it is always affected by theories, and so does not imply it cannot be used to test theories objectively. In situations in which

we carefully examine something in bright light, walk about it, etc., we seem unable to make the object look different. In such situations, our experiences can be used to objectively test theories. This does not mean, of course, the appearance of things is never deceptive under ideal conditions - perhaps a stick seen through clear and well-lit water will look bent. However, it does mean that our beliefs are not determining the way things look, so at least one source of error can be eliminated (Couvalis 1997, pp.13-14).

In sum, theory-laden perceptions have been recently discussed in philosophy of science because they seem to accredit a certain form of scepticism about scientific knowledge. The existence of theory-laden perceptions has consequently been challenged, or they have been dismissed as one of the many sources of potential error that are inherent in every experimental framework and that we can learn to control.

Now, this brief overview of the debate about the perceived threat of theory-laden perceptions to scientific knowledge is of limited interest in and of itself. What is of interest, however, is that there exists a very similar phenomenon which has not been examined by philosophers of science even though it is integral to many scientific inferences. Of even greater interest is that is it not an epistemic nuisance but rather a desirable phenomenon that can be analysed with mathematical rigour. Accordingly, I believe that we can now discuss the phenomenon of the theory-ladenness of observation in a more favourable light. This is what I propose to do in the following sections.

3 Statistical Imputation: A Well-Established Scientific Practice

It is trivial to state that data contain errors. Certain types of errors can be accounted for by our statistical models. Others can be accounted for on theoretical grounds. For example, an extreme value in our data set can be rejected because it does not make any sense according to our best theories (*e.g.*, a body temperature of 60 degrees Celsius for a living human being). In other words, we can alter our data set in conformity with our theoretical beliefs in order to account for errors in our data. In that sense our final data set can be said to be theory-laden. This kind of inferential practice has already been noted by philosophers (see Forster 1988 for example).

At this juncture, however, I wish to discuss a very different phenomenon. In some research contexts, our data are problematic not because they contain errors; the problem lies in the fact that some of them are missing, *i.e.*, our data set is incomplete. In such contexts there is no need for recourse to theoretical beliefs in order to modify our data, but we might need them in order to infer new data points. To that end, I shall now consider scenarios in which the data (bottom-up information) are complemented (not altered) by theoretical beliefs in order to arrive at scientific inferences. Such scenarios have yet to be explored in philosophy of science.

Data sets with missing data are especially common in social sciences as well as in clinical and epidemiological research. When the collection of data involves the participation of many individuals, it is likely that they will mistakenly skip questions, refuse to answer certain questions, or miss clinical appointments (Sterne et al. 2009). As a result, statisticians often have to handle missing data in order to make inferences.

There are many different methods for dealing with missing data, and the effi-

caciousness of each method depends on the type of missing data involved. Missing data are of three types. Data can be missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). This nomenclature can be found in (Rubin 1976).

In order to elucidate these notions, we must first define some terms. Let $\mathbf{X} = (X_{i1}, X_{i2}, X_{i3}, ..., X_{ip})^T$ stand for the ith observation of the p variables of a data set, where $i = (1, ..., n)$. Let j stand for the jth variable, where $j = (1, ..., p)$. Next consider $\mathbf{X}_{io}$ to be the subset of the variables that are observed for i^{th} observation and $\mathbf{X}_{im}$ to be the subset of the variables that are missing for i^{th} observation.

We must then define the variable $R_{i,j}$. It is an indicator variable that is equal to 1 if the ith observation of the jth variable is missing and equal to 0 if it is not missing. This allows us to define the missing value mechanism as follows:

$$Pr(\mathbf{R}_i | \mathbf{X}_i)$$

where

$$\mathbf{R}_i = (R_{i1}, R_{i2}, R_{i3}, ..., R_{ip})$$

If $Pr(\mathbf{R}_i | \mathbf{X}_i) = Pr(\mathbf{R}_i)$, then the missing data are MCAR. In other words, if the non-missing data are simply a random subset of the complete data set, then the missing data are MCAR. If $Pr(\mathbf{R}_i | \mathbf{X}_i) = Pr(\mathbf{R}_i | \mathbf{X}_{io})$, then the missing data are MAR. This means that the probability of obtaining a missing observation depends only on the observed data. If that probability is also dependent on the unobserved data, then the missing data are MNAR (see Carpenter and Kenward 2012).

These notions are important because they allow us to assess the use of different methods to account for missing data. For example, if our missing data are MCAR, then it might be possible to make sound statistical inferences by eliminating all the observations that contain at least one missing observation on one of the p variables. This is called a "complete case analysis". The drawback to a complete case analysis

10

when the missing data are MCAR is that the number of observations can be quite low and the power of our tests (the probability of rejecting the null hypothesis when the latter is false) can therefore also be quite low.

On the other hand, if our missing data are MAR or MNAR, then a complete case analysis can lead to biased estimates. We there need to treat them quite differently. We can do so by imputing the missing values, *i.e.*, by inferring the missing information. This is the main notion that I now wish to discuss in this paper.

In what follows, I will not review the various imputation techniques; the interested reader is invited to consult the references provided at the end of this article. My goal is to illustrate statistical imputation, explain how imputed data sets are similar to theory-laden perceptions, and thus show why theory-laden observations are created purposefully for their epistemic benefits.

To do this, I shall create a data set with MAR missing data. I shall then compare various regression estimates. Some will be computed using a complete case analysis of the data, while others will be computed using a technique called "multiple imputation" (MI). I shall underscore the benefits of the second method.

The data set with which I shall work consists of 200 observations of three variables: $v1$, $v2$, and $v3$. Here are the distributions of each variable:

$$v1_i \sim \mathcal{U}(30, 45)$$

$$v2_i \sim Binomial(1, 0.7)$$

$$v3_i \sim \mathcal{N}(2 + 0.9v1_i + 0.6v2_i, \sigma^2 = 9)$$

$$i = (1, 2, ..., 200)$$

In the following analyses, I shall take for granted that every observation for a given variable is independent and identically distributed.

In order to create MAR missing data, I draw from a binomial distribution with the parameters (1, 0.9). If I draw a 1 and if $v3 > 40$ or $v3 \leq 34$, then I remove the corresponding observation for $v1$. If not, then I retain the value for $v1$ (please consult the R code for a description of the methodology used). Overall, I remove 55% of the observations of the variable $v1$.

Now, suppose that the dependent variable is $v3$. If I were to construct a linear regression given $v2$ by ignoring all the missing data (complete case analysis) and then compare it to the linear regression that I would have obtained if I had no missing data, then I would obtain the following results:
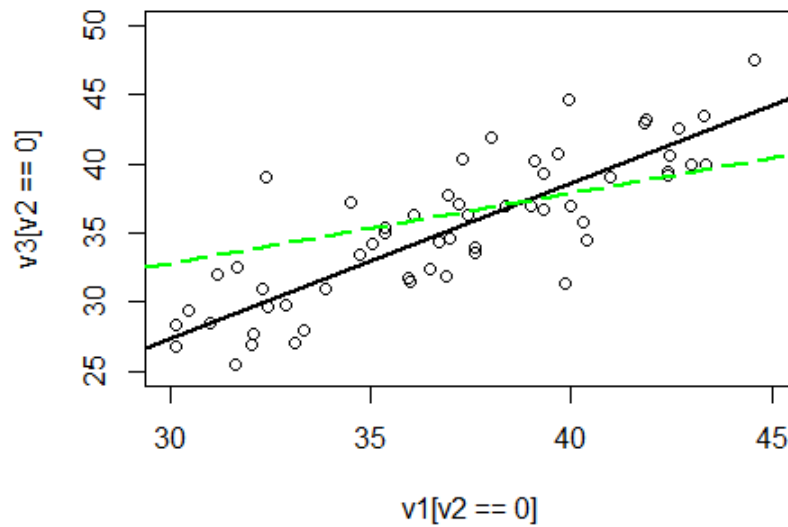

Figure 1: A linear regression estimate (black line) based on a complete data set (small circles) and a linear regression estimate (green dashed line) based on an incomplete data set. The missing data are MAR and v2=0.
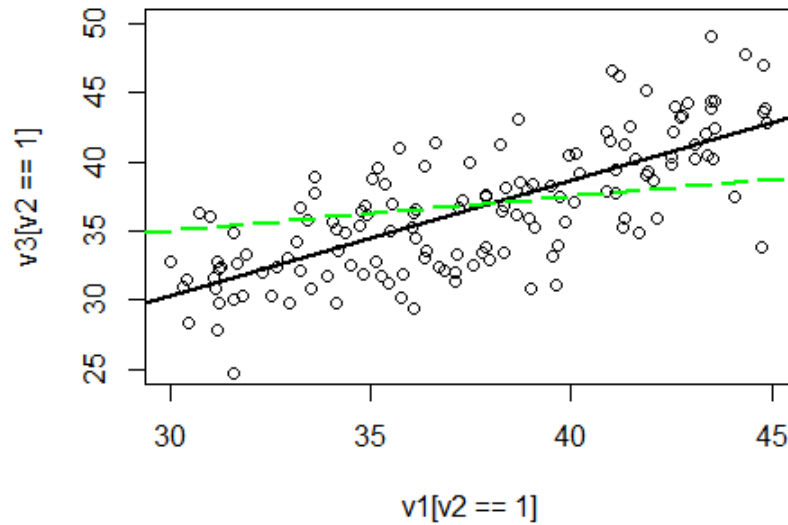
Figure 2: A linear regression estimate (black line) based on a complete data set (small circles) and a linear regression estimate (green dashed line) based on an incomplete data set. The missing data are MAR and v2=1.

The two regression lines in Figures 1 and 2 are very different. In fact, it is apparent that the regression estimates based on the incomplete data (green dashed lines) set differ significantly from the real function and also from the best estimates that we can obtain in this situation (the black lines).

We can clearly see that the missing data are a nuisance, and this is where imputed data sets come to the rescue. We shall now see that if we use multiple imputations (MI) in order to produce an estimate of the function between $v1$ and $v3$ for the two different values of $v2$, the results will be much more satisfying. MI is a technique that allows us to infer plausible values for the missing data. To be more

13

precise, we infer more than one plausible value for each missing datum in order to account for the variability of the error associated with each inference.

To perform multiple imputation in this case, I shall first use a regression model in which $v3$ and $v2$ are the independent variables and $v1$ is the dependent variable. This will allow me to replace the missing observations of $v1$ with plausible values. For each replacement value I shall add a random error (please consult the R code for details). This will create a complete data set on which I shall be able to do a regression model in which $v3$ is the dependent variable. I shall then repeat this procedure 1000 times, take an average of the regression coefficients, and thus obtain an estimate of the function between $v1$ and $v3$ for the two different values of $v2$. Here are the results that I obtained using this procedure:
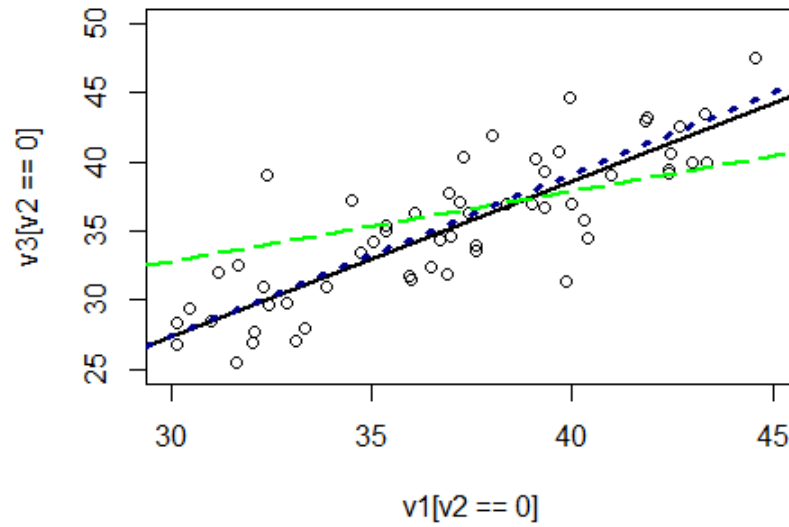
Figure 3: A linear regression estimate (black line) based on a complete data set (small circles); a linear regression estimate (blue dotted line) based on 1,000 imputed data sets; and a linear regression estimate (green dashed line) based on an incomplete data set. The missing data are MAR and v2=0.
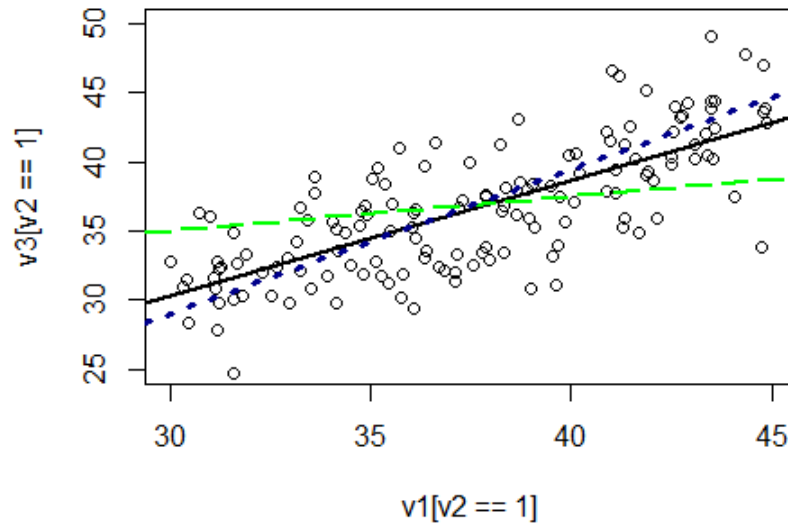
Figure 4: A linear regression estimate (black line) based on a complete data set (small circles); a linear regression estimate (blue dotted line) based on 1,000 imputed data sets; and a linear regression estimate (green dashed line) based on an incomplete data set. The missing data are MAR and v2=1.

Figures 3 and 4 show that our estimates based on multiple imputed data sets (blue dotted lines) are much closer to the best linear estimates (black lines) than those based on the incomplete data sets (green dashed lines). In other words, it is immediately evident how MI has improved the quality of our estimates.

What is more, I can take the complete data set, eliminate observations based on the MAR mechanism explained earlier, and compute a regression estimate of $v3$ in function of $v1$ and $v2$ based on the incomplete data set. If I repeat this exercise 10,000 times and take the average of the corresponding regression coefficients,

then we can form an idea of the bias caused by the missing data. If I do the same thing for a regression based on 1,000 imputed data sets, then we can also form an idea of the benefits of the imputation method. Here are the results that I obtained by doing so (please consult the R code for a description of how these simulations were performed):

Table 1: Comparison of Regression Coefficients based on 10,000 Simulations

Real Coefficients	2	0.9	0.6
Mean Value of Coefficients (Missing Data)	19.75	0.44	0.34
Mean Value of Coefficients (Imputed Data Set)	4.49	0.83	0.69

Table 1 clearly shows that we have reduced the bias of the coefficients by using MI. We may thus conclude that MI can improve our statistical inferences.

However, when we work with a real data set, we usually do not know the real function that relates our variables, we do not have access to the complete data set, and we do not know the missing value mechanism. We consequently need to use other methods to assess the benefits of multiple imputations.

In practice, we often need to rely on our knowledge of the data collection methods in order to justify the hypothesis that we are working under MCAR, MAR, or MNAR. Let us assume that the MAR hypothesis is well justified and that we need to know whether our multiple imputation has been successful. In this particular case, where it is reasonable to suppose that the observations are independent, we can use a technique called cross-validation.

One of the many ways to implement cross-validation is to remove one observation from our data set, construct a model, and then compute the square of the difference between our prediction of the removed observation and that observation. If we repeat this procedure for every observation in our data set and average the results, we will obtain a value (LOOCV) that can guide our choice of model: the smaller

17

the LOOCV the better (LOOCV stands for "Leave-One-Out Cross-Validation").
If we compute LOOCV for the different types of regression estimates that we obtained from our data set, then we observe the following results (please consult R code for details):

Table 2: Comparison of the Leave-One-Out Cross-Validation Score

LOOCV (Complete Data Set)	10.36
LOOCV (Missing Data)	16.61
LOOCV (Imputed Data Sets)	10.47

Table 2 clearly shows that the regression made with multiple imputed data sets is superior according to the LOOCV score.

We would be able to show very similar results if we had an independent data set with which to calculate a MSE score, *i.e.,* the mean square difference between the predictions of our models and the corresponding values given by the independent data set (please consult the R code for details). The results are shown in Table 3.

Table 3: Comparison of the Mean Square Error Score Calculated with an Independent Data Set

MSE (Complete Data Set)	8.65
MSE (Missing Data)	16.08
MSE (Imputed Data Sets)	8.99

In sum, whether we look at our example from a perspective in which we know the real function that relates $v3$ with $v1$ and $v2$ or from a perspective in which we are trying to assess the quality of our regression models only with the data at hand, we see that multiple imputation has clear advantages over complete case analysis. We have thus demonstrated just how helpful it can be to impute missing data and how we can assess that benefit in practice using cross-validation scores. It should

be noted that such scores are used with the observed data only, which have not been modified in any way.

4 The Comparison between Theory-Laden Perceptions and Imputed Data Sets

Looking back at the previous two sections, we can see that the imputed data sets which we have created are fairly similar to theory-laden perceptions. Both are generated from what can be called "bottom-up information" and "top-down information". In the case of theory-laden perceptions, the bottom-up information is furnished by the visual stimuli; in the case of imputed data sets, it is furnished by the observed data. When the bottom-up information is degraded, it can be improved by the top-down information.

Top-down information can be described as empirical generalisations that have been inductively inferred. These generalisations are usually referred to as "theories". In the case of theory-laden perceptions, they consist of our background beliefs about the external world. In the case of the imputed data sets that we have constructed in the previous section by using a multiple imputation technique, these generalisations are regression estimates of the function that relates the dependent variable $v1$ with the dependent variables $v2$ and $v3$. Other imputation methods can involve different generalisations.

Hence, if the expression "theory-laden perception" is adequate, then the expression "theory-laden data set" is equally so. In fact, given that data sets are referred to as "the observations," imputed data sets can also be said to be theory-laden observations. The real interest in making such a comparison, of course, is not purely semantic; it is epistemic.

Unlike in the case of theory-laden perceptions, the existence of imputed data

sets cannot be challenged. Furthermore, imputed data sets do not constitute an undesirable obstacle to scientific knowledge. As we saw in the previous section, we create them purposefully in order to improve our statistical models.

What is particularly noteworthy in the example that I have given is the fact that we need not worry about an infinite or circular regress of justification when we rely on imputed data sets. That is because cross-validation techniques allow us to assess the epistemic benefits of imputation with the observed (non-imputed) data only. Therefore, the theories used to create the imputed data set are not used to justify the model we ultimately choose, and that is one of the major epistemic differences between imputed data sets and theory-laden perceptions (we cannot compute a cross-validation score with visual stimuli).

In fact, cross-validation is possible because the theories we rely on to impute data do not alter the observed data but merely complement them. This is a particularity of imputed data sets that makes them especially interesting from an epistemological point of view, since the observed (non-imputed) data can preserve their neutrality with respect to the theories used for the imputation of the missing data. It is well known by philosophers of science that we can use our theoretical beliefs in order to alter the observed data. However, it has not yet been noted that we also use them to complement the observed data.

At last, we can now enunciate the following three theses:

1. Theory-laden observations need not be treated as an inevitable nuisance. They can be created purposefully for their epistemic benefits.

2. Certain scientific inferences cannot be understood and assessed unless we understand how and why our observations are theory-laden.

3. The theories loaded onto our observations need not play a role in the justification of our scientific models.

5 Conclusion

In this paper, I have argued that imputed data sets are analogous to theory-laden perceptions. I have accordingly shown how theory-laden observations manifest themselves in some statistical analyses. What is philosophically interesting here, however, is to realise that inferences based on imputed data sets need not be perceived as being in any way defective. On the contrary, we often purposefully load our data sets with theories in order to improve our statistical inferences, and we can obtain evidence that those inferences have been improved using empirical methods such as cross-validation.

Cross-validation allows us to assess the benefits of statistical imputation using the observed data only. This is a crucial point. It means that the theoretical assumptions which are needed to create an imputed data set do not play a role in justifying the model that we ultimately choose. The epistemic worries that are usually attributed to theory-laden observations (circular or infinite chains of justification) consequently dissolve into thin air.

I have illustrated this assertion with a simple example in which I have employed an imputation method called "multiple imputation". In doing so I have introduced important nuances into the existing philosophical literature on the theory-ladenness of observation. The latter has recently tended to focus narrowly on scepticism, relativism, and the scientific realism debate. The presentation of uncontroversial examples of theory-laden observations accompanied by an explanation of why they matter in science was overdue. This is what I have endeavored to accomplish in the present paper.

References

Bogen, J. (2014). Theory and observation in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2014 ed.).

Brewer, W. F. and B. L. Lambert (2001). The theory-ladenness of observation and the theory-ladenness of the rest of the scientific process. *Philosophy of Science*, S176–S186.

Carpenter, J. and M. Kenward (2012). *Multiple imputation and its application*. John Wiley & Sons.

Chalmers, A. F. (2013). *What is this thing called science?* Hackett Publishing.

Churchland, P. M. (1979). *Scientific realism and the plasticity of mind*. Cambridge University Press.

Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: A reply to jerry fodor. *Philosophy of Science*, 167–187.

Couvalis, G. (1997). *The philosophy of science: science and objectivity*. Sage.

Feyerabend, P. (1993). *Against method*. Verso.

Fodor, J. (1984). Observation reconsidered. *Philosophy of Science*, 23–43.

Fodor, J. A. (1991). The dogma that didn't bark (a fragment of a naturalized epistemology). *Mind*, 201–220.

Forster, M. R. (1988). Unification, explanation, and the composition of causes in newtonian mechanics. *Studies In History and Philosophy of Science Part A 19*(1), 55–101.

Gillies, D. (1993). *Philosophy of science in the twentieth century: Four central themes*. Blackwell.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 352*(1358), 1121–1127.

Hanson, N. R. (1958). Patterns of discovery an inquiry into the conceptual foundations of science.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago press.

Little, R. J. and D. B. Rubin (1989). The analysis of social science data with missing values. *Sociological Methods & Research 18*(2-3), 292–326.

Maxwell, G. (2009). The ontological status of theoretical entities. *Philosophy of Science: An Historical Anthology*, 451.

McCauley, R. N. and J. Henrich (2006). Susceptibility to the müller-lyer illusion, theory-neutral observation, and the diachronic penetrability of the visual input system. *Philosophical Psychology 19*(1), 79–101.

Papineau, D. (1979). *Theory and meaning*. Clarendon Press.

Raftopoulos, A. (2001). Reentrant neural pathways and the theory-ladenness of perception. *Philosophy of Science*, S187–S199.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association 91*(434), 473–489.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, Volume 81. John Wiley & Sons.

Sterne, J. A., I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood, and J. R. Carpenter (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj 338*, b2393.