Collaborative practice, epistemic dependence and opacity: the case of space telescope data processing

Julie Jebeile

Abstract

Wagenknecht recently introduced a conceptual (yet non-exhausting) distinction between translucent and opaque epistemic dependence in order to better describe the diversity of the relations of epistemic dependence experienced by scientists in collaborative research practice. In line with her analysis, I will further elaborate the different kinds of expertise that are specific to instrument- and computer-assisted practices, and the potential sources of opacity. For that, I focus on a contemporary case of scientific knowledge creation, i.e., space telescope data processing.

Résumé

Wagenknecht a récemment introduit une distinction conceptuelle (non exhaustive) entre dépendance épistémique translucide et dépendance épistémique opaque, dans le but de mieux rendre compte de la diversité des relations de dépendance épistémique au sein des pratiques collaboratives de recherche. Dans la continuité de son travail, mon but est d'expliciter les différents types d'expertise requis lorsque sont employés instruments et ordinateurs dans la production de connaissance, et d'identifier des sources potentielles d'opacité. Mon analyse s'appuie sur un cas contemporain de création de connaissance scientifique, à savoir le traitement de données astrophysiques.

1. Introduction

In contemporary science, knowledge is increasingly the outcome of collective enterprises that divide epistemic labor, and, therefore, members come to depend upon one another epistemically. Epistemic dependence in collaborative research practice is thus now gaining attention in social epistemology (e.g., de Ridder 2014, Andersen 2016, Andersen and Wagenknecht 2013, Wagenknecht 2014; after the major work of Hardwig 1985). While discussions on this matter most often focus on the evidential status of testimony and the nature of the knowing subject, Wagenknecht recently (2014) refined the concept. She introduced a conceptual (yet non-exhausting) distinction between translucent and opaque epistemic dependence. As she defines it, "A scientist is opaquely dependent upon a colleague's labor, if she does not possess the expertise necessary to independently carry out, and to profoundly assess, the piece of scientific labor her colleague is contributing. [...] if the scientist does possess the necessary expertise, then her dependence would not be opaque, but translucent" (2014, p. 483). Such a distinction aims at better describing the diversity of epistemic dependence scientists experience in research practice; it situates two extreme poles of epistemic dependence between which a variety of other instances lies.

This paper is an original contribution to Wagenknecht's conceptual analysis. I will further elaborate the different kinds of expertise that collaborative research requires, and the different potential sources of opacity. Science's escalating reliance on supporting technologies, including instruments and computers, will be emphasized, alongside, which will become clear, these technologies' determination of the ways and means of epistemic dependence. Therefore, the analysis will be of social epistemology, and partly of philosophy of science in practice as well.

More precisely, I will apply Wagenknecht's conceptual distinction on a contemporary case of scientific knowledge creation, i.e., space telescope data processing. Both human and material organizations process the data and operate the telescope from the Earth's surface. This case exemplifies today's collaborative research practice for at least three reasons: firstly, it is part of a large-scale project funded by several governments; secondly, it (thereby) requires international scientific teams to work together; and, thirdly, it relies strongly on interdisciplinary collaborations. Astrophysical images are now often built from space telescope data whose measure and digital processing involve specifically competent multinational teams of astrophysicists, i.e., instrumentalists, experts in computer programming and specialists in data analysis. Here each collaborating scientist depends crucially on the other contributions being adequately conducted and truthfully reported.

I will identify cases of opaque epistemic dependence within the human and material organization behind the whole process. In space telescope data processing, astrophysical images are produced by a group of scientists following a series of sequential tasks. Data processing is a sequential chain of data transformation, from the raw data measured by the telescope to the processed data from which scientific analysis can actually start and knowledge be visually inferred. Each step of the process depends upon the integrity of the preceding step. Each part is performed by a particular individual or a specific epistemic subgroup who therefore depends upon one another epistemically.[1]

2. Translucent vs. opaque epistemic dependence

Beforehand, I need to introduce Wagenknecht's concept of epistemic dependence and the distinction she offers between translucent and opaque epistemic dependence. In doing so, I will highlight that the main features of epistemic dependence, the way she defines it, are suitable for collaborative research, and that her polar distinction between translucent and opaque epistemic dependence properly describes the diverse relationships experienced by contemporary scientists.

First, contrary to Audi (1983), Wagenknecht's concept of epistemic dependence is not about static belief-belief relations (that concern the justification of one belief based on another) but rather about *dynamic relations between scientists*. As she defines it, "When a scientist's knowing, i.e., her believing-that and/or the reason for her believing to be justified, crucially involves what her colleagues know or the epistemic efforts that they have undertaken, then we have a case of epistemic dependence" (Wagenknecht 2014, p. 477-478). Her concept thus stands at the level of dependence relations between people (rather than of epistemic relations between beliefs) which

[1] Other configurations of division of epistemic labor, in which scientific tasks are parallel or can be made parallel, are therefore excluded from this analysis; they may nevertheless be less problematic with regard to opaque epistemic dependence.

better characterize the way epistemic labor is actually delegated in collaborative research: in practice, beliefs are held by individuals, and these individuals interact with each other.

Second, epistemic dependence arises from asymmetries in epistemically relevant resources, including most importantly the *expertise* distributed among team members; other resources comprise particular experimental devices, time, 'sheer' labor, social capital, etc. Epistemic dependence is "between one scientist's *knowing* on the one hand and another scientist's *expertise*, as manifest in his knowing and his epistemic labor on the other hand" (*ibid.*, p. 477, emphasis mine). Here expertise is meant to be built from "manual and cognitive resources, both declarable and tacit, such as experience and technical skills that are acquired through a process of professional maturation" (*ibid.*, p. 476). Asymmetries in expertise are due to epistemic labor: only some team members have to conduct experiments, and therefore will eyewitness them and take measurements; only some are in charge of the data analysis and will draw inferences from the data. That said, asymmetries in expertise can also be due to the fact that research incorporates people from different disciplines (in an interdisciplinary team), or different specialties (in a mono-disciplinary team), who therefore possess diverse scientific backgrounds.

Third, Wagenknecht emphasizes the material dimension of scientific practice in order to grasp an important property of epistemic labor: physical devices are constantly used to provide experimental evidence, which therefore "is mostly conveyed in a material form, i.e., as a photograph, a copy of the laboratory notebook, a print, a frozen sample, a concrete model, a diagram, or a file" (*ibid.*, p. 478).

A significant consequence of this property, for Wagenknecht, is that while a scientist relies on an experimental work produced by her colleague, she needs not necessarily rely on his believing that this work is reliable. As soon as the material results of the experiments performed by her colleague are available to her, and under the condition that she possesses the required expertise to interpret them, then she can form a belief that *p* based on them and she can make her own judgment as to whether the epistemic labor of her colleague is evidence of *p* or not. This is a case of translucent epistemic dependence as we will later see.[2] Such an instance of epistemic dependence is not a simple belief–belief relation.

Once she has identified the possibility for a scientist to make her own judgment on another's work, Wagenknecht is in a position to argue for a refinement of the concept of epistemic dependence in collaborative research practice. To start with, a scientist is not epistemically dependent when she has, on her own, *sufficient* evidence for why a statement 'that *p*' should be regarded as reliable. Having sufficient evidence for *p* means (i) to actually perform the piece of scientific labor (e.g., to conduct or scrupulously eyewitness a measurement), as well as (ii) to possess the necessary expertise to perform it and understand the procedure. The original contribution of Wagenknecht is to take into account conceptually that, in practice, without having performed the piece of scientific labor herself, the dependent scientist may have some or full expertise on the issue in question, and may be in a position to acquire on her own a *part* of (but not all) the required evidence. From this, she draws her distinction between opaque and translucent epistemic dependence, which aims at describing the diversity of instances of epistemic dependence in collaborative research practice.

When the dependent scientist has *full* expertise on the issue at stake, then we have a case of translucent epistemic dependence. She relies on someone else for pragmatic reasons, often related to considerations of time, but because she possesses expertise pertaining to the claim that

[2] In order to form a belief that *p* based on the material results of the experiments performed by her colleague, the scientist needs to assume that these experiments have been correctly performed. The material results could indeed be themselves misleading. That is why here epistemic dependence cannot be qualified as transparent, but can still be qualified as translucent.

p, she can at least partially assess her colleague's claiming that *p* – and, if available, she can assess the experimental evidence that her colleague provides for this claim.

When, by contrast, the dependent scientist does *not* have *any* expertise pertaining to the piece of epistemic labor at stake, then we have a case of *opaque* epistemic dependence. The dependent scientist has no means to establish the truth that *p* other than assuming that her colleague is an honest, skillful testifier and/or a competent expert. This is the kind of dependence that epistemologists usually talk about (e.g., Hardwig 1985).

When the dependent scientist has *some* expertise on the matter at stake, then we are in the grey zone. Within this zone, there are clearly gradations of more and less expertise.

Now that I have presented Wagenknecht's conceptual distinction, I will focus on a contemporary case of scientific knowledge creation, i.e., space telescope data processing. My first objective is, based on this example, to elaborate further on the different kinds of expertise that are required in collaborative research practice, especially those involved in instrument- and computer-assisted practices.

3. Space telescope data processing: the case of Herschel

Distribution of scientific labor in collaborative practices, and formation of collective knowledge by groups of scientists, have received widespread attention in science studies after the practical turn (e.g., Knorr-Cetina 1999), especially when related to Big Science (e.g., Price 1963 ; Galison 1987, 1997). Here, a particular contemporary case is studied: today most astrophysical images are not photographs (McCray 2014). Their production goes from the measure of raw data by a telescope (in a space mission or in a terrestrial observatory) to their reduction. This entire process involves a social (and international) organization, i.e., the ground segment, made of competent teams of astrophysicists, including instrumentalists, experts in computer programming and specialists in data analysis, and supported by relevant instruments and computers. How can we ensure that, at the end of this social process, the astrophysical images are reliable?

No one could in practice possess all the required kinds of expertise for conducting or re-conducting each step of the process, and for having control over the entire process; a division of labor is required for obvious pragmatic reasons. The scientists who are designing the next space mission, Euclid, and received feedback from the space mission Herschel, wrote: "Many individuals, scientists and engineers, are and will be involved in the [Euclid Scientific Ground Segment] development and operations. The distributed nature of the data processing and of the collaborative software development, the data volume of the overall data set, and the needed accuracy of the results are the main challenges expected in the design and implementation of the Euclid [Scientific Ground Segment]." (Pasian et al. 2012, abstract).

At the very best, we could expect to have translucent epistemic dependence each time epistemic dependence is required. In such an ideal situation, each member has access to the data as well as sufficient expertise to be able to assess the parts of the process on which her work depends. Opaque epistemic dependence is the least desirable: if errors occur within her input, the dependent scientist would still rely on them and the errors will therefore propagate into the rest of the process, endangering the reliability of the astrophysical images. In this section, my attention will focus on the data transformation—or the "data journeys", i.e., the circulation of data across widely different research contexts and locations (Leonelli 2015)—within the Herschel mission. Based on this example, I will map, within the telescope data processing, the areas of opaque epistemic dependence.[3]

A difference with Wagenknecht's examples here is that the scale of my case study is not a scientific laboratory (as she takes the cases of a molecular biology laboratory and a planetary science group). This is rather a particular research program, which requires important financial investments and appropriate infrastructure, like any of Big Science's collaborative project (e.g., Human Genome Project, Human Brain Project, Laser Interferometer Gravitational-Wave Observatory, The Large Hadron Collider), and (thereby) involves several international laboratories.

That the case study is a particular research program makes the question of opaque epistemic dependence more dramatic. On the one hand, the program is based on a series of sequential tasks, each of them performed by an agent (or a subgroup), and each of them depending on one another's outputs. Therefore, if an error occurs at one step, it may impact the rest of the series. On the other hand, a program is conducted on a limited period of time, and this authorizes less moving back and forth between contributors for solving identified issues (when they are identified).

The use of digital data processing is a salient aspect of contemporary scientific imaging, "in addition to the instrument, imaging devices also rely on computers, not just as display devices but also as machines that permit to store and retrieve data as well as to transform them, mathematically" (Israël-Jost 2016). In particular, instruments and computers are nowadays both required in the production of astrophysical images which starts with the measure of raw data by a telescope and goes on with a series of data reductions; the whole process involves a background social organization. This will be now illustrated by the case of Herschel Space Observatory, which was launched in 2009. Such a case study is new in the philosophy of science.

Let me first briefly present the instrumental part of the Herschel data production. Because of its large single mirror, Herschel Space Observatory has been used to collect long-wavelength radiation from some of the coldest and most distant objects in the Universe. The observatory is composed of (i) the telescope, which is a set of mirrors converging the light to the instrument, (ii) the instruments, and (iii) the service module which contains all the electronic part dedicated to the command of the satellite and the instrumentation.[4] The instruments, located in the satellite, detect light and measure the raw space data, which constitute the "source images". There are three instruments within the satellite, i.e., HIFI, PACS and SPIRE. For example, PACS is composed of detectors, spectrometers and photometers. Light passes through the PACS entrance optics and is simultaneously transmitted via the mirror to the spectrometers and the photometers. It is then converted into electronic signals through an instrumental chain of detections (made of optics, detectors, reading system, communication system) (for more details, see Morin et al. 2004). The measured data—called the raw data—convey information about relevant physical magnitudes, such as wavelengths, velocities of objects in space, and densities. They are temporally stored in the computer of the satellite, and only in due time sent back to the ground in the form of wave packets.

The observational and measuring systems also heavily rely on computers. Once transmitted to the ground, the data must then be algorithmically processed so to be corrected from several sources of errors, notably from measurement artifacts. Going from raw data to processed data is about making successive data reductions. A data reduction is a transformation of raw data collected by the instruments into a more organized and simplified form (see Hoeppe 2014 for an anthropologist description of data reduction and data analysis). The algorithmic procedures for data reduction are described in the extraction pipelines. The processed data are the completion of the space missions

[3] This case study benefited from discussions with astrophysicists Vera Könyves and Marc Sauvage during my immersion in 2015/2016 at the Astrophysics Department, CEA Saclay.
[4] See http://herschel.esac.esa.int/Docs/Herschel/html/ch02.html.

as, from them, images can be built, made available to the scientific community, and used to draw information about the properties, morphologies and compositions of the objects in space under study.

I will now highlight that the whole data processing relies on an important social organization, i.e., the ground segment. First the ground segment plays an essential role by launching the observatory and by providing support in setting up and in operating the instrument for acquiring raw space data. Second it carries out the whole process of producing processed data from raw data.

As shown in Figure 1, the Herschel Science Ground Segment (H-SGS) consists of six elements: the Mission Operations Centre (MOC), the Herschel Science Centre (HSC), the NASA Herschel Science Center (NHSC), and three Instrument Control Centre (ICCs), one for each Herschel instrument (HIFI, PACS, SPIRE).

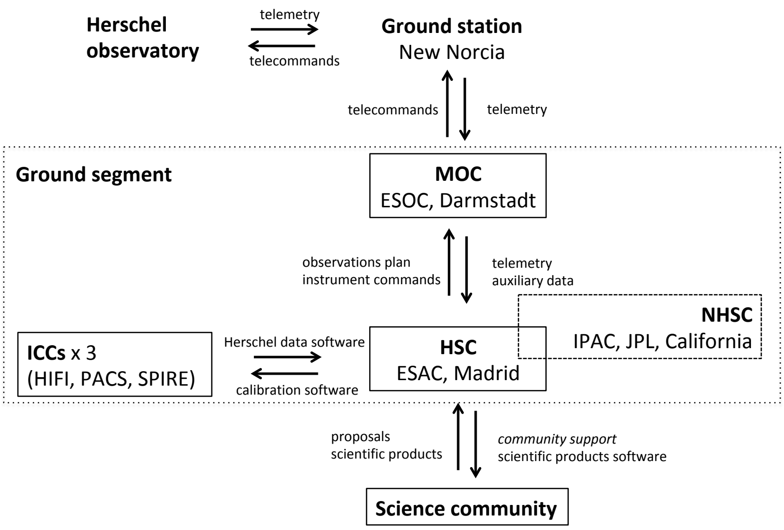

Figure 1. Herschel organization

Each element plays a specific role: The MOC sets up the instrument into the satellite, prepares the planning of observations, and sends the observation requests to the observatory. It also receives the measured data. The ICCs perform instrument monitoring and instrument calibration, and they develop instrument specific software for observations and data processing. The HSC provides to the science community information and user support related to the entire life-cycle of Herschel observations. It provides the catalogs of data and delivers the extraction pipelines. It also performs scientific mission planning, produces observing schedules and provides this information to the MOC. The NHSC provides additional user support for the US-based users.

The described elements all take part in the data processing in a sequential way. The MOC performs the first treatments on the raw data, then sends them to the HSC on a daily basis as well as to the ICCs for quick-look analysis. The data undergo a first generic reduction by the HSC, then are stored in the local database at the HSC and made available to the scientific community.[5] They undergo a second reduction by the ICCS following a multi-level process (that will be later presented). The scientific community performs an additional specific reduction on the data, before using them for data analysis. Reduction here depends on the instrument used as well as the kind of questions raised. Data analysis is then done with the help of visualization tools and post-processing analysis.

In view of this organization, it becomes clear that the cooperation between the different actors, including astrophysicists, dynamics and systems engineers, attitude control and measurement

[5] See for example http://gouldbelt-herschel.cea.fr/archives/.

hardware designers, star-tracker manufacturers, is important to attain the expected quality of astronomical images, while all the contributors have their own scientific specializations and come from institutions located in different countries. In what follows, I will highlight the different kinds of expertise required for the data processing that are specific to instrument- and computer-assisted practices.

Beforehand, I briefly want to point out that the ground segment is constrained by institutional and political considerations, such as the necessity of attributing operations to international centers at different geographical places: the ground station is based at New Marcia, Australia, the MOC at ESOC, Darmstadt, the HSC at ESAC, Madrid, the NHSC at IPAC, Pasadena, and the ICCs are provided by their respective principal investigators. Thus the organization satisfies political choices which depend more on the history of international institutions involved financially and scientifically in the space mission (Krige and Russo 2000; Krige, Russo and Sebesta 2000). In that sense, it does not strictly satisfy an epistemic logic, and therefore very unlikely does it optimize the distribution of cognitive resources and the network of epistemic dependency relations. This may be an important aspect for this study, but I will not examine it here.

4. Expertise in collaborative instrument- and computer-assisted practices

Wagenknecht insists on the gradual nature of epistemic dependence: starting from opaque, epistemic dependence may become grey and even translucent. For her (2014), the dependence is translucent if the dependent scientist has the expertise to assess the evidential and argumentative pieces for *p*, before using *p* in her own piece of work. It is opaque if the dependent scientist does not have such a situated expertise, for she did not get the time and/or the opportunity, throughout her career or in her current job, to develop the relevant skills and competences. Such expertise can thus be gained over professional experience, by learning to master experimental techniques independently for instance. Therefore epistemic dependence can be made less and less opaque. Situated expertise may nevertheless be hard to acquire over time. This is particularly true within interdisciplinary groups where opacity is attributable to group members' different fields of expertise while jointly contributing to research questions.

While, for Wagenknecht, the possibility for a scientist to gain expertise (and therefore for epistemic dependence to become less opaque) is restricted by lack of time and/or opportunity, I want now to claim that such a possibility may also be restricted by the very kind of objects to which required expertise in collaborative research practices pertains. In Herschel, like in many other cases (see Collins 1985/2002; Collins and Evans 2007), collaborative research practices involve the use of instruments and computers. Instruments and computers therefore partly determine the division of cognitive labor and the ways and means of epistemic dependence. I will now explicate the required pieces of expertise before presenting in the next section their corresponding sources of opacity.

In the example, for the processed data to be reliable, pieces of expertise are required about (i) the preparation of observations, (ii) the calibration, functioning and use of the instruments for acquiring raw data, and (iii) the design and implementation of the pipelines for reducing the data.

(i) First of all, from the start, two distinct kinds of knowledge in physics and in computer science are required in order to prepare the observations. Expectations of what is physically relevant and should be observed by the telescope are then translated into mere computational terms. The problem then becomes one of applied mathematics. Thus, before the launch, for the relevant data to be measured, a selection of expected measurements is prepared in the Astronomical Observation Requests by the HSC. The most interesting star-forming sites to be scanned by the telescope are chosen by the group of physicists, in an approximate manner (see, for example, the selection of the denser regions to be scanned in the nearby molecular clouds, in Aquila and Serpens' regions, made by the Herschel Gould Belt survey consortium: http://www.herschel.fr/Images/astImg/23/gb_serpens.jpg).

Then, in the Herschel Observation Planning Tool (Hspot), observations are planned in more detail, using the parameters that Herschel actually needs for carrying out observations. These parameters are, for example, the center coordinates of the scanned boxes, the scanning direction and the speed of the detector on the sky, the instruments (SPIRE and PACS) and its mode (photometric map, or spectroscopy). Preferentially the instruments are used simultaneously to scan the selected regions, thus allowing to obtain a map which gives a homogeneous wavelength coverage for the ranges of interest. The scan lines by SPIRE cover the wavelengths of 250-350-500 micron, and the scan lines by PACS cover the wavelengths of 70-160 micron. This way, the targets under study can be later reconstructed during data analysis.

(ii) For the instrumental part, expertise in calibration, telemetry and verification of basic instrumental functions is required in order to set up the instrument into the satellite, and ensure instrument monitoring and calibration. Physics of the target, i.e., long-wavelength radiation from the coldest and most distant objects in the Universe, is required to identify the physical phenomena, properties, morphologies and compositions of the objects in space under study that are observed. The theoretical principles underlying the instrument's functioning, i.e., the optics underlying the spectrometers and the photometers' functioning, must also be known, notably to anticipate optical artifacts within observations on telescope, aberrations, deformations, and the overlapping of objects. Basic knowledge in electronics and in computer programming may also be important in order to treat the data once they are converted into electronic signals, or compressed to be stored in a computer.

(iii) For the computer part, expertise in physics and mathematics are needed in order to assess the content of the pipeline that describes the algorithmic procedures to reduce the data; the pipeline helps in correcting the data from relevant sources of errors such as measurement artifacts. In the data processing, transforming tasks, leading to data of different levels, are roughly subdivided into merely mathematical or physical transformations. Here is the overview of Herschel Data Products Levels as it is defined by scientists working on the project. It shows at which level mathematics and physics considerations are taken into account:

Level-0 data products are raw telemetry data as measured by the instrument. They might be minimally formatted before its ingestion into the *Herschel Science Archive* (HSA). They are automatically generated by the data processing pipeline.

Level-1 data products are detector readouts calibrated and converted to physical units. They are in principle independent from the theory of the instrument and the functioning of the observatory.

Level-2 data products are obtained after level-1 data be further processed to such a level that scientific analysis can be performed. (See an illustration of level-2 data products for the region of Aquila in (Bontemps et al. 2010, Figure 2) and (Könyves et al. 2010); the residual scan lines indicate that an additional treatment is needed to make pictures more readable).

In a nutshell, *level-2.5 and level-3 data products* are data measured for different wavelengths, and then combined together. In this way, relevant photometric and spectroscopic pictures of the objects under study can be reconstructed. (See an illustration of level-2.5 and level-3 data products for Aquila Pictures in (Könyves et al. 2015, Figure 1).

The scientist should also be able to assess whether the unfolding computational processes of the pipeline lead to reliable outputs and should therefore know the epistemic criteria of verification and validation used in such a context as well.

5. Sources of opacity in data processing

Now that I have explicated the kinds of expertise that are specific to instrument- and computer-assisted practices, like the ones in the Herschel space mission, I will identify instances of opaque epistemic dependence within the human and material organization behind the whole process.

Wagenknecht's work stepped to a practical turn in social epistemology as she endeavors to enrich her conceptual analysis with relevant details concerning scientific collaborative practices. This analysis will go a step further by considering how social aspects of research and chosen technologies can actually influence epistemic dependence relations among scientists. Collaborative practices in science include more than mere epistemic exchanges. More precisely, I will show that (i) social aspects of research, such as competition and industrial secrecy, can make epistemic dependence opaque, and (ii) inherent aspects of the used technological (i.e., instrumental and computational) support can determine epistemic dependence relations.

5.1. Non-disclosure of data

First of all, as already said, a precondition for translucent epistemic dependence is that the evidential and argumentative bits and pieces for *p* are made available. And yet, while open data often stand as a *communist* ideal norm in science (Strevens, *forth.*), there are in practice reasons of institutional and political kinds for not disclosing data. When collaborators are part of a same project for which their respective institutions have given money, competition is expected to be avoided and thereby access to data restricted. Basic norms of *coopetition*—a blend of cooperation and competition—are thus established, especially now that making accessible and sharing data are easily done online: "[…] the digitization of astronomy helped reshape norms and behaviors in the astronomy community—what scholars describe as a 'moral economy'" (McCray 2014), "In terms of scientific research, a *moral economy* refers to the often-unstated rules, values, expectations, and obligations associated with the production of knowledge and circulation of resources" (*ibid.*).

In the Herschel mission, each team has access to a certain set of data that is temporarily restricted to the other teams. For planning observations, the Guaranteed-Time Key Programme is elaborated in advance, based on a constitution,[6] in order to distribute observing time to each team of a same Consortium. Herschel counts three Consortia—one for each instrument SPIRE, PACS, HIFI. For instance, in return for the time spent in the development of the system, the SPIRE consortium received a guaranteed 10% of the Herschel mission observing time. This time has then been divided among the six specialist astronomy groups, which are part of this consortium. Each group focuses on a different area of scientific investigation—e.g., observations of high redshift galaxies, local galaxies, or targets within our own galaxy and high-mass star-forming regions beyond. Each group has about ten to twenty months in order to exploit the data for itself. Thus, within a consortium, for this period at least, scientists cannot compete. Competition is implicitly authorized only after all the data being finally stored in HSC and made accessible online to the whole scientific community all over the world (in virtual observatories[7]).

Even if data are made available, there is still a certain form of opacity since, in space missions, data are produced in mass. For example, Euclid, which is the forthcoming ESA mission for mapping the geometry of the dark Universe, will deliver roughly 850 Gbit of compressed data per day in all likelihood. Thus the difficulty of treating and analyzing huge quantities of data appears.

5.2. Industrial property

[6] See for example http://www.herschel.fr/cea/gouldbelt/en/Phocea/Vie_des_labos/Ast/ast_visu.php?id_ast=23.
[7] For example, the archive IRSA , NASA/IPAC: http://irsa.ipac.caltech.edu/frontpage/ ; or Herschel science archive: www.cosmos.esa.int/web/herschel/science-archive.

Lack of knowledge about the functioning of the instrument is also a possible source of opacity. Instrumental processes are sometimes hidden to the scientists for financial/industrial reasons (e.g., industrial property), and therefore constitute black boxes. This is an instance of opaque epistemic dependence (without collaboration) between scientists and industrial furnishers.

An example of industrial property that created opaque epistemic dependence in Herschel is the pointing accuracy in the star-tracker (Schmidt and Salt 2012; Sánchez-Portal et al. 2014). Verification of its in-orbit performance revealed inaccuracy in the alignment of the different fields of view of the instruments with respect to the telescope. The pointing accuracy was therefore monitored and improved by the scientists themselves during various campaigns as the furnisher refused to indicate the manufacturer specifications of its functioning because of industrial property. Herschel was then providing observations with accuracy in the order below one arcsec for the Absolute Pointing Error.

5.3. Epistemic opacity of numerical calculations

In the data processing, numerical calculations are run for reducing data. Here the required expertise are at least about (i) the content of the computer program, i.e., the pipeline, and (ii) the unfolding computational processes of the machine. And yet these two objects are not always accessible to the agent.

First, the scientist may learn the content of the program because she conceived and/or wrote and/or carefully read the program. But actually users of a computer program generally do not have the occasion to carefully read the content of the program insofar as users and developers are typically divided into two distinct groups. Developers mainly focus on the implementation, i.e., the writing of the algorithms, and on the verification of the computer program while users execute the program and post-process the outputs. Thus, users might not have access to all the details in the computer program.

Second, the scientist should be able to assess the unfolding computational processes of the machine based on her memory and cognitive capacities (see Burge 1998, Barberousse and Vorms 2014). And yet, because of the cognitive limits of human beings, this is a highly difficult task. One of the reasons is that these processes are epistemically opaque. About the opacity of computer simulations, Humphreys writes:

> In many computer simulations, the dynamic relationship between the initial and final states of the core simulation is epistemically opaque because most steps in the process are not open to direct inspection and verification. This opacity can result in a loss of understanding because in most traditional static models our understanding is based upon the ability to decompose the process between model inputs and outputs into modular steps, each of which is methodologically acceptable both individually and in combination with the others. (2004, p. 148)

Opacity thus prevents the user from comprehending the whole detailed process by which the simulation produces the connection between the program and the outputs. One source is that simulations run so fast that no human brain could follow or survey the computational processes in detail. For this same reason, computationally assisted proofs of mathematical theorems, such as the four-color theorem, are often controversial (McEvoy 2008). Even in the case where their speed was reduced and adapted to the cognitive skills of the user so that she could follow the simulation unfolding,[8] she would need a lot of time, due to the large number of calculation steps, to follow the

[8] I borrow this notion from Dubucs (2006) who characterizes derivation in formal systems "as the process of unfolding the mathematical content of the axioms by means of the progressive application of the inference rules". He adds that "running a computer program can be viewed as unfurling the content implicit in its instructions."

calculation entirely, and she would be unable to cognitively grasp it anyway. Each computational step is understandable, but it is not possible to master the calculation in extenso.

Opaque epistemic dependence makes it difficult for the dependent scientist to trace the origin of an error in the whole process—in calibration? measurement? reduction?—(and maybe even identify the error itself) and straightforwardly suspect the dubious data.

6. Robustness of the data processing

In the division of cognitive labor in data processing, human resources are rationalized in that the pieces of scientific labor, given their epistemic nature (i.e., mathematical, physical or merely computational), are assigned to agents having the relevant kinds of expertise (i.e., in mathematics, physics or computer science). Given the sources of opacity, what does scientists' trust in the whole data processing then build on? My answer is that such a division of labor is legitimated when robustness of the processing is met. And, robustness is measured when back and forth exchanges between collaborating scientists stop.

A common assumption about collaborative practices in science is that epistemic dependence is based on testimony (Hardwig 1985). An implicit idea is that the testimony of a scientist to another is expressed once. This should nevertheless be de-idealized in cases of scientific collaborations. For such cases, one should take into account that exchanges of evidence and arguments between two collaborators working in the same research project are often not straightforward, but are rather based on a back-and-forth communicational dynamics between the two, which creates some inertia in the collaborative activities of the group. The dependent scientist can ask for more clarification or information, and sometimes for revisions. This is an important aspect in actual practices, as it is frequent that exchanged data can appear to be dubious to the dependent scientist in case she obtains unexpected or inaccurate results from them.

I suggest that back-and-forth exchanges, and subsequent corrections in the process, stop when the whole computer-assisted instrumental process is considered to be robust (or stable). Instrumental outputs are traditionally taken to be robust by scientists if (i) "[the instrument] fulfills, reliably and durably, the function for which it has been conceived (the measurement of this or that quantity or phenomenon)"; and (ii) "the instrumental outputs […] provide information about the 'true values' of the variables, about characteristics actually possessed by the measured objects" (Soler 2012, p. 16). The same should hold for computer-assisted instruments generating algorithmically processed data. Once the group agrees to consider the process robust, then the process is expected to lead to reliable empirical data for data analysis to start.

In the general case, robustness of an instrument allows the division of labor between, on one side, the instrumentalists who know the theory of the instrument, and, on the other, the scientists who exploit the instrumental outputs. Hacking famously claimed that "One needs theory to make a microscope. You do not need theory to use one" (1981). For example, knowing how a microscope works is not required for the biologist who uses it to interpret her observations, e.g. what she sees as microbes. (Whether investigators must know and be able to use the theory of the instrument to correctly interpret the outcome has been nevertheless widely discussed in the literature, and Hacking's position has been roundly criticized.)

In the case study, the ground segment must ensure that all the operations in the data processing be properly conducted, and control the continuity between them. There has nevertheless been some back and forth within the process (see Pasian et al. 2012), meaning that the processing has not been robust from the beginning. At the end of the Herschel mission, scientists have long identified, on the maps, errors from data reduction, revised the pipelines accordingly, and redone the calculations. According to astrophysicist Vera Könyves with whom I was in contact, depending on the final Herschel product, stabilization has been reached after two to four years. This might seem low compared to the time the data will be actually exploited, i.e., roughly several decades, but still high given the important competition for publications. Once robustness is reached, revisions stop, as well as exchanges about them, and the final processed data can be released to the scientific community and be exploited for first scientific publications. In the example, reduction of back-and-forth exchanges have been accelerated by the role of a mediator, Vera Könyves, who was at the intersection of all the required kinds of expertise and was in charge of anticipating requests from the users to the developers.

From then on, once robustness is met, opaque epistemic dependence should become less problematic. Users are supposed to be able to draw knowledge from the final processed data without having to know the theory of the instrument. Once the final Herschel products have been proven to be robust, astrophysicists have been able to more trustworthily conduct data analysis without knowing how first-levels data have been actually obtained, i.e., without knowing the details of data reductions.

7. Conclusion

In this paper, I have applied Wagenknecht's concepts of translucent and opaque epistemic dependence on a contemporary case of scientific knowledge creation, i.e., space telescope data processing taking place on a ground segment. My objective was to identify cases of opaque epistemic dependence within the human organization behind the whole process. After having explicated the kinds of expertise required throughout this process, I have shown that there are other sources of opacity than the lack of expertise that Wagenknecht indicates, i.e., non-disclosure of data, industrial property and the epistemic opacity of numerical calculations. I have finally suggested that robustness of the processing may be an indication of it trustworthiness.

Remerciements : je suis reconnaissante aux deux évaluateurs·trices anonymes de cet article pour leurs précieux conseils. Je remercie également le Service d'Astrophysique du CEA Saclay de m'avoir fourni les conditions à la fois matérielles et intellectuelles à l'étude de la mission spatiale Herschel. En particulier, je remercie vivement les astrophysiciens Vera Könyves et Marc Sauvage pour nos discussions à ce sujet.

References

Andersen, H. (2016) Collaboration, interdisciplinarity, and the epistemology of contemporary science. *Studies in History and Philosophy of Science* 56 1-10

Andersen, H. and Wagenknecht, S. (2013) Epistemic dependence in interdisciplinary groups. *Synthese* 190:1881–1898

Audi, R. (1983) Foundationalism, Epistemic Dependence, and Defeasibility. *Synthese*, 55: 119–39.
Hardwig, J. (1985) Epistemic Dependence. *Journal of Philosophy,* 82 .7.: 335–49.

Barberousse, A. and Vorms, M. (2014) About the warrants of computer-based empirical knowledge. *Synthese.* 191:3595–3620.

Bontemps, S., André, P., Könyves, V. et al. (2010) The Herschel first look at protostars in the Aquila Rift. *Astronomy & Astrophysics*, 518, L85.

Burge, T. (1998). Computer Proof, Apriori Knowledge, and Other Minds: The Sixth Philosophical Perspectives Lecture. *Noûs*, 32 (Supplement: Philosophical Perspectives, 12, Language Mind and Ontology):1–37.

Collins, H. (1985, 2002) *Changing Order: Replication and Induction in Scientific Practices*. Chicago: University of Chicago Press.

Collins, H. and Evans, R. 2007. *Rethinking Expertise*. Chicago, IL: University of Chicago Press.

de Ridder, J. (2014) Epistemic dependence and collective scientific knowledge. *Synthese*. 191:37–53

Dubucs, J. (2006). Unfolding Cognitive Capacities. In M. Okada (Ed.) *Reasoning and cognition*, (pp. 95–101). Keio: Keio University Press.

Galison, P., (1987) *How Experiments End*. University of Chicago Press, London.

Galison, P. (1997) I*mage & logic: A material culture of microphysics*. University of Chicago Press.

Hacking, I. (1981) Do We See Through a Microscope?, *Pacific Philosophical Quarterly* 62: 305-22.

Hardwig, J. (1985) Epistemic Dependence. *Journal of Philosophy* 82 .7.: 335–49.

Hoeppe, G. (2014) Working data together: The accountability and reflexivity of digital astronomical practice. *Social Studies of Science*, Vol. 44(2) 243–270.

Humphreys, P. (2004) *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford: Oxford University Press.

Israel-Jost, V. (2016) Computer Image Processing: An Epistemological Aid in Scientific Investigation. *Perspectives on Science*, vol. 24 no. 6, 2016, pp. 669-695.

Knorr Cetina, K. (1999) *Epistemic Cultures: How the Sciences Make Knowledge*. 1st ed., Harvard University Press.

Könyves, V., André, P., Men'shchikov, A., et al. (2010) The Aquila prestellar core population revealed by Herschel, *Astronomy & Astrophysics*, 518, L106.

Könyves, V., André, P., Men'shchikov, A., et al. (2015) A census of dense cores in the Aquila cloud complex: SPIRE/PACS observations from the Herschel Gould Belt survey, *Astronomy & Astrophysics*, 584, A91.

Krige, J. and Russo, A. (2000) *A History of the European Space Agency 1958 – 1987. Volume I. The story of ESRO and ELDO, 1958 - 1973.* Available online on the website of the European Space Agency.

Krige, J., Russo, A. and Sebesta, L. (2000) A History of the European Space Agency 1958 – 1987. Volume II. The story of ESA, 1973 to 1987. Available online on the website of the European Space Agency.

Leonelli, S. (2015) *A centric biology-a philosophical study*. Chicago: University of Chicago Press.
McCray, W. P. (2014) How Astronomers Digitized the Sky. *Technology and Culture*, Volume 55, Number 4, pp. 908-944.

McEvoy, M. (2008). The Epistemological Status of Computer-Assisted Proofs. *Philosophia Mathematica*, 16(3):374–387.

Morin, B., Okumura, K., Sauvage, M., Gastaud, R., Andreani, P. M., Panuzzo, P., Turek, G. (2004) The PACS Simulator. Report v. 1.6.6, 2004/09/30, document of CEA/DAPNIA/SAp Saclay. ftp://ftp.mpe.mpg.de/pub/ftp_ir/sturm/GSR/PACS-SIM-Saclay-02.pdf

Pasian, F., Hoar, J, Sauvage, M., Dabin, C., Poncet, M., Mansutti, O. (2012) Science ground segment for the ESA Euclid Mission. *Proc. SPIE 8451, Software and Cyberinfrastructure for Astronomy II*, 845104

Price, Derek J. de Solla (1963) *Little Science, Big Science ... and Beyond*. Columbia University Press.

Sánchez-Portal, M. et al. (2014) The Pointing System of the Herschel Space Observatory. Description, Calibration, Performance and Improvements. https://arxiv.org/pdf/1405.3186.pdf

Schmidt, M. and Salt, D. (2012) Herschel Pointing Accuracy Improvement. Hitting the bullseye from 3234,03 meters distance - History of the Herschel Attitude Control System Pointing Accuracy Improvement. http://www.spaceops2012.org/proceedings/documents/id1275622-Paper-001.pdf

Soler, L. (2012) Introduction: The solidity or scientific achievements: structure of the problem, difficulties philosophical implications. In L. Soler, E. Trizio, T. Nickles and W. C. Wimsatt, *Characterizing the Robustness of Science. After the Practice Turn in Philosophy of Science*. Boston Studies in the Philosophy of Science, Springer

Strevens, M. (forthcoming) Scientic Sharing: Communism and the Social Contract. Forthcoming in Thomas Boyer-Kassem, Conor Mayo-Wilson and Michael Weisberg (eds.) Scientific Collaboration and Collective Knowledge, Oxford University Press

Wagenknecht, S. (2014) Opaque and translucent epistemic dependence in collaborative scientific practice. *Episteme* 11 (4):475-492