European Journal for Philosophy of Science (2021) 11: 10
https://doi.org/10.1007/s13194-020-00325-6

PAPER IN PHILOSOPHY OF SCIENCE IN PRACTICE

IBE in engineering science - the case of malfunction explanation

Kristian González Barman[1] · Dingmar van Eck[1,2]

Received: 19 November 2019 / Accepted: 16 October 2020 / Published online: 6 November 2020

Abstract

In this paper we investigate how inference to the best explanation (IBE) works in engineering science, focussing on the context of malfunction explanation. While IBE has gotten a lot of attention in the philosophy of science literature, few, if any, philosophical work has focussed on IBE in engineering science practice. We first show that IBE in engineering science has a similar structure as IBE in other scientific domains in the sense that in both settings IBE hinges on the weighing of explanatory virtues. We then proceed to show that, due to the intimate connection between explanation and redesign in engineering science, there is a further engineering domain-specific virtue in terms of which engineering malfunction explanations are evaluated, viz. the virtue of *redesign utility*. This virtue entails that the explanatory information offered by a malfunction explanation should be instrumental in predicting counterfactual dependencies in redesigned systems. We illustrate and elaborate these points in terms of a number of engineering examples, focussing in particular on the 2009 crash of Air France Flight 447. Our extension of analyses of IBE and explanation to engineering science practice offers new insights by identifying a new explanatory virtue in malfunction explanation: redesign utility.

Keywords Abduction · Diagnosis · Engineering science · Explanatory virtues · Failure analysis · Inference to the best explanation · Redesign utility

✉ Kristian González Barman
KristianCampbell.GonzalezBarman@Ugent.be

Dingmar van Eck
Dingmar.vanEck@Ugent.be; d.vaneck@uva.nl

[1] Centre for Logic and Philosophy of Science, Department of Philosophy and Moral Sciences, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium

[2] Department of Philosophy & Institute for Logic, Language, and Computation (ILLC), University of Amsterdam, Amsterdam, Netherlands

1 Introduction

Eventually all artefacts will break down, but some do so too early or unexpectedly. The analysis and prevention of such malfunctions calls for engineering disciplines like reliability engineering, failure analysis, and engineering forensics. In such disciplines, the prevention of future malfunctions often hinges on the explanation of current or past malfunctions; e.g., the explanation that the 2010 BP oil spill was caused by a malfunctioning blowout preventer was a main driver of redesign efforts aimed at constructing better functioning ones. When such malfunction explanations are sought by analysts, they want the best one possible out of the available alternatives, relative to specific goals (and constraints), such as improved designs that enable preventing future failures, or explanations that enable adjudicating legal responsibility. Yet, despite the fact that Inference to the Best Explanation (IBE) is central to many diagnostic tasks in engineering science, there appears to be no philosophical account on offer of how IBE works in this domain, neither in philosophy of science nor in philosophy of technology. Our aim in this paper is to assess how IBE works in the engineering practice of explaining technical malfunctions.

Indeed, the lack of such an account of IBE is quite puzzling given that many authors agree that IBE is fundamental to diagnostic reasoning, both in medical and technological contexts (Lipton 2004; Josephson and Josephson 1994; Bhaumik 2009; Schupbach 2017). So, while IBE, inter alia, has been analysed in the context of scientific methodology (Lipton 2004; Psillos 2002), legal decision making (Amaya 2007), testimony (Fricker 2017), and medical diagnosis (Bird 2010; Dragulinescu 2016, 2017), the area of technology remains unexplored.

In-depth study of IBE in engineering malfunction analysis also extends the literature on technical functions in philosophy of technology. In accounts of technical function, definitions and justifications of malfunction ascriptions often accompany characterizations and justifications of (normal or proper) function ascriptions[1] (Preston 1998; Krohs 2009). The ICE theory of technical functions (Houkes and Vermaas 2010), for instance, lists malfunction ascription as one of the main desiderata that a theory of technical artifacts should be able to specify. But this literature has not addressed how IBE is carried out in engineering practice. Similarly, examination of IBE in engineering malfunction analysis extends extant work on (malfunction) explanation in engineering science (van Eck and Weber 2014; Van Eck and Weber 2017; van Eck 2015, 2016); while this work addressed the (mechanistic) structure of malfunction explanation, the reasoning steps involved in the weighing and selection of explanations in engineering practice remained unaddressed. In this paper we aim to elaborate this aspect of the engineering practice of explaining (and redesigning) malfunctioning technical systems.

The main aim of this paper is to examine the use of *inference to the best explanation* (IBE) in failure analysis by extending current IBE literature to engineering science. Since IBE hinges on judging explanatory virtues of explanations, our *first objective* is to chart some prominent virtues that are involved in the evaluation of explanations in

[1] We are not concerned here with the philosophical project of spelling out what makes function and malfunction ascriptions *justifiable*. Such normative accounts abound in philosophical theories of function, but are irrelevant for our purposes here, viz. elaborating how IBE works in the engineering *practice* of explaining malfunctions.

failure analysis. This provides a theoretical apparatus to achieve our *second objective*, which is to spell out how these virtues relate to one another and how they figure in the evaluation of the goodness of malfunction explanations. We address these issues in terms of a number of engineering examples, focussing in particular on the 2009 crash of Air France Flight 447. Through these two objectives we reach our main aim. Our *third objective* is to compare IBE in malfunction explanation in engineering science with IBE in other scientific domains.[2] A new insight that we derive from this comparison is that there is an engineering domain-specific virtue in terms of which engineering malfunction explanations are evaluated, viz. the virtue of *redesign utility*. This virtue entails that the explanatory information offered by a malfunction explanation should be instrumental in predicting counterfactual dependencies in redesigned systems.

We start Section 2 with some conceptual groundwork, clarifying key notions that are present in our analysis of IBE in engineering practice: malfunction, malfunction explanation, IBE, and explanatory virtue. Section 3 elaborates some salient virtues operative in engineering diagnostic reasoning (first objective). Section 4 clarifies how these virtues interact in IBE in engineering practice, using AirFrance-flight-447 as a case study (second objective). Section 5 compares IBE in malfunction explanation in engineering science with IBE in other scientific domains (third objective). Section 6 offers conclusions.

2 IBE, malfunction, and explanatory virtues

This section clarifies key notions that are present in our analysis of IBE in engineering practice: malfunction, malfunction explanation, IBE, and explanatory virtue. We address these notions in turn.

2.1 Malfunction

The term *malfunction* can be understood in many ways in engineering practice. Del Frate (de Frate 2014: 47–51) lists no less than 54 different definitions, including deficits, errors, anomalies, breakdowns, and failures. We do not aim to flesh out all the subtleties around the meanings of the term malfunction but rather use the generic notion that something malfunctions when "the item stops performing its required function" (Birolini 2007:3), in which we take the 'something' to either be an entity, activity, or organizational feature of a technical system. But we stress that our project is not a definitional one; we merely use this notion since it allows us to regiment engineering practices of explaining (and preventing) malfunction. Since our project concerns elucidating reasoning steps involved in engineering practice by which engineers evaluate and choose explanations of malfunctioning *technical* systems, it excludes several other projects. So, for instance, analyses according to which engineers,

[2] We use the terms 'engineering science', 'engineering', 'engineering practice', and 'engineering science practice' interchangeably in this paper. All research practices of engineers that we refer to by these terms we consider to be *scientific* practices. This paper is thus about engineering science, not about engineering: we take *engineering science* to refer to research that is aimed at developing new technical artifacts or at improving them; *engineering* in turn refers to the activity of constructing technical artifacts, based on knowledge gathered by engineering scientists.

rather than technical systems, fail given that proper designs would have prevented the malfunctions (Feld and Carper 1997) fall outside the scope of our analysis. Similarly, analyses that question whether a malfunctioning 'X' still counts as an 'X' (Jespersen and Carrara 2013) are also orthogonal to our purposes and not considered here. Our analyses do however include cases in which the interaction between user actions and malfunctioning (sub)systems is what caused an accident.

The above definition of course presupposes normal conditions of use (e.g. the inability to drive a car across a lake is, by any reasonable standard, not a malfunction). Also, accidents are not malfunctions, yet within a diagnostic context it can be challenging to prize the two apart (a crash could have been due to severed brakes, or the breaks could have severed because of the crash). Furthermore, there is a distinction between token and type failure that can roughly be identified as the difference between an item not conforming with the design – production issues, improper transportation, etc. –, and the design itself being the problem. From this point of view, we could explain a malfunction through 3 different contrasts: a difference between the artefact and its design, between the design and an improved design (which would not create problems), or between normal conditions and conditions that can cause a failure. We will focus on explanations that consider the issue to be one of design, and which aim at improving current designs.

A final comment on terminology and scope. In the engineering literature, an observation of a malfunction is often called a *symptom* and an explanation of a malfunction often called a *diagnosis*. Some malfunctions are predictable and their diagnosis (and solution) is predetermined; e.g. repair instructions in a manual or the practice of Failure Mode and Effect Analysis (FMEAs), where the task is to try to prevent as many malfunctions as possible beforehand. But some malfunctions and failures are hard or impossible to predict and require a posterior diagnosis.[3]

2.2 Why questions and explanation types

When attempting to explain a malfunction, it is important to clearly define the why-question (with its relevant contrast class). A malfunction can be explained in a variety of ways depending on what is being asked. For example, both a short-circuit and a lack of security checks may explain the same malfunction of a smartphone.[4] ('why did Ana's Samsung phone rather than her Samsung laptop explode?' vs 'why did Ana's Samsung phone rather than her iPhone explode?'). Furthermore, the characterisation of the contrast class also affects the level of abstraction at which an explanation is pitched (e.g., a cooker whose left and right stoves are turned on by the wrong switches does not malfunction at a higher level, since it does afford cooking, but it malfunctions at a lower level, since turning the right switch does not activate the right cooker).

[3] Plumbridge (2009) studies the case of "Tin whiskers". When changing led for a Tin alloy (for environmental reasons), it created nanofibers (whiskers). Sometimes two whiskers from nearby plaques connect, creating a short-circuit. This problem could not have been predicted, since the problems that arise at smaller levels are not analogous to problems at a macro-level.

[4] For instance, Yun et al. (2018) studied how a closed innovation strategy and unidirectional chain of command led to batteries' malfunctioning in Samsung Galaxy Notes 7 in 2016, due to which the smart phones ignited in a number of cases, whereas Loveridge et al. (2018) delved deep into the chemistry and physics involved in the short-circuiting batteries.

Additionally, one can construct nested explanations by turning the explanans of an explanation into a new explanandum for another explanation; some companies implement this technique, iterating it up to 5 times, to get the appropriate depth of explanation – a procedure known as the '5 whys', which was popularized in the 50s by Toyota.[5] (We here endorse Woodward and Hitchcock's (2003) notion of *depth* according to which the greater the range of counterfactual inferences an explanation affords making, the more *depth* it has).

In failure analysis, most explanations are arguably either causal or causal-mechanical. Such explanations answer the contrastive question 'why malfunction, rather than normal or expected function?' and articulate factors – failing components or (sub) mechanisms – that are taken to make a difference to the occurrence of a specific malfunction (van Eck 2016; Van Eck and Weber 2017).[6]

Several techniques are in use in engineering to reconstruct the interaction of these factors, such as Fault Tree Analysis that weighs causal factors (deductively) to give the likelihood of a failure, and Root Cause Analysis that tries to separate root causes and causal contributing factors (usually employing causal graphs). As said, which features are identified is relative to the why question asked and the abstraction level at which the explanation thus is pitched. In this paper we focus on causal-mechanical malfunction explanations.

2.3 IBE

The terms 'abduction' and 'inference to the best explanation' (IBE) are often mixed in the literature, but it is important to clarify the differences for our purposes. Abduction refers to a non-monotonic inference type aimed at formulating a hypothesis given an observation, such that if it were true, it would explain the observation. In contrast, IBE is an argument that (often) uses an abductive pattern to justify the validity of an explanation. IBE can take many forms, but it usually entails three premises: a fact, a disjunctive list of possible explanations, and the affirmation that one explanation is better than the others.

There are, in general, two approaches for studying IBE. One approach investigates the structure or *format* of inference patterns in IBE arguments (Niiniluoto 1999; Schurz 2008; Hoffman 2010; Bird 2010), which most authors consider to be abductive. The other attempts to spell out what justifies an explanation as the best. Such justification is often procured by citing *explanatory virtues* (Lipton 2004; Thagard 1978; Psillos 2002). The best explanation is then considered to be the one that satisfies a given set of explanatory virtues optimally. However, some of these virtues might come into conflict, as Niiniluoto (1999) has shown.

Our analysis fits the second approach: examining and weighing the 'goodness' of explanations vis-à-vis competitors. In some cases, this goodness might be based entirely on evidence alone. This is what Bird (2010) considers to be the case in some

[5] See, e.g. Serrat, Olivier (2017).

[6] There are a number of different proposals in the literature for understanding mechanistic explanations. We endorse a generic notion of mechanistic explanation that most parties in the debate will agree on: mechanistic explanations articulate the mechanisms, organized collections of entities and activities, that produce, underlie, or constitute explanandum phenomena (Machamer et al. 2000; Bechtel and Abrahamson 2005; Craver 2007; Glennan 2005, 2017; Illari and Williamson 2012).

medical (differential) diagnoses, where one considers all possible explanations, and then eliminates by evidentiary virtues, leaving only one. But, oftentimes, evidence alone is not sufficient to unequivocally decide in favour of one explanation over others (e.g., evidence might be scarce, might point in several directions, or might be interpreted in different ways): there are cases where we cannot eliminate all options. Another issue when relying on evidence alone is that one may not have good positive reasons to believe that the explanation selected indeed is a good one. This signals the need to invoke other criteria for the ranking of explanations: explanatory virtues. In the cases we analyse both evidence and explanatory virtues play important roles.

While in the cases we analyse, explanations have a causal or causal-mechanical format, IBE is not restricted to specific explanatory formats: one can rank different explanations having the same format, e.g., causal-mechanical as in our case studies, as well as ones having different formats, e.g., causal-mechanical and inductive-statistical. Furthermore, IBE can be investigated at different levels, e.g., inferences of individual scientists, teams of scientists, fields of scientists, and can include a temporal dimension, in which different 'rounds' of IBE are investigated over time. These features surface in our case study of the 2009 crash of Air France Flight 447: different explanations for this crash have been inferred as best at different points in time, both individually and collectively.

2.4 Explanatory virtues

Different types of virtues have been distinguished in the literature. For our current purposes, the distinction between evidentiary or evidential virtues[7] (those that are linked to evidence) and explanatory virtues (those that are linked to explanation and understanding[8]) is important. In our examples and cases, explanatory virtues are crucial in IBE and we hence focus on this type of virtue. Explanatory virtues name desired characteristics or properties of theories or explanations.

A great variety of explanatory virtues have been analysed in the literature, such as *consilience* (accounting for more data with less hypotheses), *simplicity* (theories and explanations should not specify superfluous or unessential hypotheses or entities), *analogy* (connections of theories and explanations to already established mechanisms or concepts) (Thagard 1978); *completeness* (accounting for most of the observations), *importance* (accounting for the most salient observations), *parsimony* (semantic simplicity), *unification* (subsuming more observations under the same solutions) (Psillos 2002); *non-sensitivity* (robustness to changes in initial or normal conditions), *accuracy* (percentage of falsehoods or idealizations), *degree of integration* (relationship to already known theories) (Ylikoski and Kuorikoski 2010); and *depth* (the range of counterfactual inferences an explanation affords making) (Woodward and Hitchcock 2003; Weslake 2010).

[7] Psillos (2002) argues that in the process of IBE, evidence is already taken into account, i.e., in the inference to a best explanation, evidence and its quality has already been weighed and considered.

[8] We use the notion of understanding advanced by Ylikoski and Kuorikoski (2010), where understanding is taken to be an inferential ability over counterfactual situations: the ability to answer contrastive *what-if-things-had-been-different questions* relating values of the *explanans* to values of the *explanandum,* where answering more (and more important) counterfactuals implies better understanding. Having a better understanding will allow for better manipulations and, in an engineering context, better redesigns.

Some of those virtues are also prominent in diagnostics in engineering science; we focus on *simplicity*, *coherence*, *depth* and *completeness*, and elaborate how these virtues are understood in engineering malfunction explanations by reference to a number of engineering examples of failure analysis. By this we achieve our first objective. We then proceed to show in section 4 how these virtues interact and function in IBE, i.e., in the weighing and selecting of explanations. Our main example there is the 2009 crash of Air France Flight 447. Section 4 achieves our second objective. Section 5 achieves our third objective by comparing IBE in engineering failure analysis with IBE in other scientific domains.

3 Explanatory virtues of malfunction explanations

As mentioned, some of the virtues discussed in the previous section are also used in IBE in engineering science. In this section we elaborate how *simplicity*, *coherence*, *depth* and *completeness* are invoked in engineering diagnostic reasoning.

Simplicity is also endorsed in engineering failure analysis. Imagine two bulbs that are connected in parallel and, even though the circuit is closed, neither of them emit light. There are at least two competing explanations: 'main supply is not giving electricity' and 'both lights are broken'. Snooke and Price (2012) in this situation posit that all else equal we should prefer the simpler explanation (main supply is broken): "Parsimonious principles dictate that the diagnosis f3 [the simpler explanation] would be a preferred initial diagnosis." (883).[9] Such principles are operative when it is not possible to perform a simple test to choose amongst competing explanations. Simplicity is a constraint that prevents us from citing too many, possibly irrelevant, causal factors, but it also has a stronger reading: out of two prima facie equally explanatory sets of (possible) causal factors, we should prefer the smaller set. The first interpretation relates to the ergonomics of understanding (repetition or irrelevant details are dispensable), but the second interpretation has to do with competing explanations and is based on the idea that the less assumptions we make the less committed our answer is.

Coherence with background knowledge of technical systems is also relevant in the selection of diagnoses (Affonso 2007; Boukharouba et al. 2009). As Snooke and Price (2012) tell us: "an engineer would make these selections and create conditional symptoms based on extensive system knowledge" (Snooke and Price 2012: 875) In other words, the strength of an explanation is affected by its *coherence* with technical

[9] This light bulb example is used as a running example to demonstrate how an automated method for failure analysis outputs malfunction explanations in the domain of (unmanned) autonomous aircraft in which these explanations are also connected to redesign issues (Snooke and Price 2012). The explanations generated by the method for instance are used to determine "whether all system failures can be observed with the existing sensors, and indicates which component failures can be isolated with existing sensors" (p. 871). In other words, explanations may reveal that a redesign of sensors is required. In line with this, part of the method is a model-generated FMEA by which "the consequences of every failure are reported to the designers, and they can decide the steps needed to improve the safety and reliability of the system." (p. 871). If, say, an explanation for malfunctioning light bulbs indicates that the wiring connecting the light bulbs of a warning system to an aircrafts' sub system for measuring altitude is prone to fracture or the switches in the circuitry are prone to corrosion, this may drive a redesign of parts of the circuitry – e.g., a different spatial configuration of the wiring or a novel type of switch. Redesign features as a component in the automated method for failure analysis from which this light bulb example derives.

background knowledge. Take for instance the explanation that the independent company *Instrumental* offered for the ignition of a number of Samsung Galaxy Notes 7, before Samsung[10] gave its reading of the incidents: "Any battery engineer will tell you that it's necessary to leave some percentage of ceiling above the battery, 10% is a rough rule-of-thumb, and over time the battery will expand into that space. Our two-month old unit had no ceiling: the battery and adhesive was 5.2 mm thick, resting in a 5.2 mm deep pocket. There should have been a 0.5 mm ceiling. This is what mechanical engineers call line-to-line — and since it breaks such a basic rule, it must have been intentional." (2016). *Instrumental's* argument is based on coherence with standard technical background knowledge and practice, and how Samsung deviated from what was known to be acceptable.

Such background knowledge is composed of analogies with other cases, models, and other things such as physical laws that set the conditions of possibilities (e.g. they allow to compare the energy of a chemical reaction to the pressure a certain section of material can withstand). In this sense, *coherence* also works by transferring explanatory power from other disciplines such as physics. If certain (accepted) equations can show why, given certain parameters, the event had to happen, this is perfectly explanatory (in virtue of coherence with these laws). These equations provide counterfactuals: they tell us something about what would happen if the conditions were different (e.g. if the variables changed).

Clustered inside this notion of coherence we find *analogy* and *unification*. Thanks to background knowledge one can make analogies with other relevantly similar cases. Furthermore, such background knowledge informs applications of a diagnosis to other problems or problem domains – we can use the same reasoning patterns, equations and information of causal factors for diagnosing other malfunctions. This is key to FMEAs — e.g., Samsung's battery crisis produced an 8-point security checklist for batteries that is widely endorsed.

Coherence therefore makes an explanation better in virtue of making more inferences possible (providing more understanding), but at the same time, the more connections there are between the established knowledge and the explanation, the more confidence it warrants.

Depth is also a virtue operative in engineering failure analysis (Affonso 2007). The depth of an explanation refers to the range of counterfactual inferences an explanation affords making (Woodward and Hitchcock 2003). Most diagnoses are causal, so the more causally relevant factors identified in a diagnosis, the better (all else equal), for it allows the tracking of more counterfactual dependencies, which in turn allows for better informed future prevention of malfunctions and more informed repair measures (see also section 5). To see how this works, consider Samsung's explanation (2017) of the first series of malfunctioning batteries in which a deflection of the negative electrode and the location of the tip in the curve area were cited as causal factors, and Samsung's explanation of the second series of malfunctioning batteries in which penetration of the

[10] During the summer of 2016 a number of Samsung Galaxy Note 7 phones were catching on fire. Samsung's engineers promptly concluded that this had to do with faulty SDI batteries produced by a subsidiary supplier (Samsung SDI co.). However, the new supplier also started having issues with the batteries shortly after. Samsung hence declared: "We recognised that we did not correctly identify the issue the first time and remain committed to finding the root cause." (2018: 271). By the 23th of January 2017 they gave a press conference where they explained what had happened together with several independent organizations.

separator, due to welding burrs, was cited as the causal factor plus, in the case of some batteries, the lack of insulation tape. Now, the independent company UL that was hired by Samsung to also provide an analysis, gave deeper explanations, combining multiple causal factors. The causal factors cited in the malfunction explanation for the first series of batteries was as follows: "A combination of deformation at the upper corners + thin separator + repeating mechanical stresses due to cycling, causing higher possibility of separator damage leading to an ISC between aluminium and copper foil at the corner." (2017:6). The causal factors cited in the malfunction explanation for the second series of batteries was as follows: "The combination of (1) missing insulation tape + (2) sharp edged protrusions on tab + (3) thin separator, all leading to a high possibility of an ISC between cathode tab and anode, subsequently resulting in heating and fire" (12). Since these explanations articulate more causal factors, they enable the tracking of more (and more varied) counterfactual dependencies.

Completeness, finally, is also valued in engineering failure analysis: explanations that account for more symptoms are considered better. As Snooke and Price put it: "it is desirable to detect as many faults as possible in as many operating states as possible" (Snooke and Price 2012: 875). In the simple light bulb example mentioned earlier, an explanation that only cites one of the bulbs as broken leaves unexplained the symptom of the other bulb failing to emit light. An explanation that explains the malfunctioning of both bulbs accounts for more of the observed symptoms and hence is more complete than one accounting for only one bulb malfunctioning. Additionally, it is not only accounting for more symptoms that is better, but so is accounting for the most relevant symptoms (what symptom belongs to a *malfunction* and which to a *defective* device is a pragmatic issue that depends on what the main function is considered to be).

For the most part, these virtues are synergistic: the more coherence and depth an explanation has, the more completeness it will arguably have. However, there exists a trade-off between simplicity and the virtues of depth and completeness (the more causal factors we cite, the deeper an explanation becomes, but it also loses simplicity).

We have illustrated with a number of brief examples that a variety of explanatory virtues, as analysed in the philosophy of science literature, are also endorsed in engineering diagnostic practices. This explication of explanatory virtues is our theoretical apparatus that we apply in the next section to clarify how these (and some other) virtues play a role and interact in IBE in engineering malfunction analysis.

4 IBE in engineering malfunction analysis: The case of air France flight 447

The 2009 Air France Flight 447 crash is – while a grave catastrophe– an intriguing case, because the final analysis of what happened refers to a variety of factors that affected the crash – bad weather conditions, malfunctioning components and, according to several analysts, a faulty design. While official reports (from the French civil aviation authority, the *Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile* (BEA), and the Judiciary report) differ on the precise transcription of the black boxes, the causes of the crash are now pretty well understood. Furthermore, interestingly, these were mostly understood *before* the black box was recovered, showing the value of IBE in the absence of evidence. In this section we argue that investigations of

the crash can be understood and regimented as a collective endeavour focused on finding the best explanation for the crash, in the process of which different explanatory hypotheses are put forward and evaluated at different points in time and, finally, one has been inferred as the best. The final explanation advanced by the community of fault analysts was to a large extent confirmed when the black box was recovered. We argue that explanatory virtues played an important role in the evaluation and selection of the different explanatory hypotheses offered.

On the first of June of 2009 Air France Flight 447 (Airbus 300–203) vanished from the sky without leaving a trace or a mayday call. It had left from Rio de Janeiro to Paris, and as it approached the Intertropical Convergence zone (which has extreme weather conditions and is outside the radar scope of any country), it simply vanished, after sending several automatic signals of electronic circuit malfunction.

Some early voices, such as Air-France spokesman François Brousse suggested thunder strike as the most plausible explanation. John Hansman (MIT aeronautics and astronautics) in an interview for CNN also defended this possibility. They hypothesized that, upon entering a raging storm with a fly-by-wire control, thunder could have struck the aircraft, disabling the system and the aircraft's backup computers (provided that the components were not well grounded).

At the time, all that was known was that the plane had entered a turbulent zone and had sent 24 automatic fault messages via satellite (including one reporting on a loss of air pressure). This information gave rise to speculations about either extremely heavy turbulence or thunder coupled with turbulence. Others suggested an electrical fault, where the fly-by-wire control would override the pilot's decisions. This claim was strengthened based on *coherence* with knowledge of past examples (e.g. Australian-Qantas-A330 where inertia sensors overrode manual commands, overcorrecting by pointing the nose down). Similarly, some suggested that this possible electrical fault might have subsequently caused a fire on the plane, just as had happened in a Swiss-Air Boeing jet in 1998.[11]

These possible early stage explanations fit the evidence that the debris, meteorology reports and the fault signals provided, but none captured all the symptoms in *complete* fashion and all lacked *depth* as well compared with successor explanations (as advanced by, e.g., the BBC/Nova investigation team and the French civil aviation authority; more on this below). The lightning hypothesis was considered the least strong, because of the lack of *coherence* with background knowledge: many planes get struck by lightning every year (albeit not in such extreme weather conditions) without resulting in crashes, and the last plane that crashed because of lightning occurred more than 40 years ago. However, these initial explanations hinge on *simplicity* so as to compensate for lack of knowledge. After recovery and analysis of the debris, analysts acquired some more information. The plane had crashed (in one

[11] Some other speculations were offered, but quickly disregarded because of their implausibility. For instance, a hail storm that would have caused a broken windshield (but then the pilots would have reported distress); a terrorist attack, which would explain the mysterious disappearance without a mayday call (but had there been an explosion, they would not have identified so many bodies in the debris). Highjack was also ruled out given the lack of radio warning and missile attack was disregarded as a possible explanation because of the altitude at which the plane flew. These explanatory hypotheses were discarded largely on evidential grounds.

piece) into the ocean, with its nose pointing up (at a 5° angle). This points towards a stall,[12] raising the question why the plane had stalled. The messages it sent also indicated that autopilot and engine thrust had been disengaged, and that the plane was flying in alternate law (where the computer does not balance the pilot's actions). These features disproved the lightning hypothesis further and required novel explanatory hypotheses to account for these symptoms, i.e., explanations that are more *complete* and have more *depth*.

One of the 24 failure messages contained information on the basis of which one could explain the disengagement of autopilot and engine thrust: the pitot tubes had malfunctioned. Pitot tubes measure aircrafts' airspeed, which is crucial for autopilot to control the aircraft. If these sensors are malfunctioning, autopilot disengages, followed by the engine thrust and the conversion to alternate law, because the computer cannot operate safely without speed indications. While, after several months of gathering and interpreting evidence (the pieces of the plane recovered, meteorology reports and the 24 automatic fault messages), the BEA did not want to make any final claims, several experts (e.g., the now retired Air France pilot Gérard Arnoux, president of the *Comité de veille de la sécurité aérienne* (aviation safety watch committee), who currently works as a technical advisor in criminal court cases) concluded there was a pretty good picture of what had happened (Traufetter 2010).

Investigations continued. The BBC and Nova (Darlow Smith, Scott 2010) composed an independent team of engineers, pilots, and meteorologists to investigate the crash and subsequently made a documentary detailing a possible explanation for it (before the recovery of the black boxes) that was quite accurate in retrospect (after recovery of the black boxes).[13] The explanation offered factors affecting why the plane flew into a storm, why the pitot tubes had malfunctioned, and why it had stalled. It was thus more *complete* than predecessor explanations. We take these why questions in turn and, by reconstructing the reasoning steps of the team involved in procuring answers to them, show that the explanations offered for them, again, resulted from IBE reasoning.

The first question they attempted to answer is why the plane flew into a storm (against standard procedure), which had not been part of the BEA's first reports. The team reasoned that the most likely explanation for this is that a smaller weather disturbance blocked the radar from detecting the actual threat, i.e., the storm. This is a *simpler* explanation than one citing technical malfunction of the radar as a cause, and also more *coherent* than assuming the pilots flew intentionally into the storm, since such an action goes against any pilot's knowledge and protocol.

[12] An aerodynamic stall occurs when the wings do not produce enough lift, either because of insufficient speed or a wrong angle of attack. This is a very dangerous situation where the plane loses altitude quickly.

[13] The explanations provided by the BBC and NOVA team are bona fide engineering explanations. The BBC and NOVA team were led by accredited experts and thus reflect engineering thinking and expertise on malfunction explanation. These experts included engineering failure analysts, aviation safety consultants, and aerospace and aviation meteorologists (see e.g., Williams et al. (2009) in which Air France Flight 447 is also investigated or Craig et al. (2008) on the detection of in-cloud turbulence). The documentary itself is referenced in the books "The Rio/Paris Crash: Uncovering the Secrets that Changed Aviation History" (Rapoport 2011) and "Understanding Air France 447" (Palmer 2013). The latter recovers a few threads of the documentary, making certain things more precise, such as the fact that supercooled water droplets likely adhered to snow crystals (forming graupels); and also giving further evidence to support this claim.

The team then proceeded to clarify what happened in the storm itself.[14] This investigation led them to explain why the pitot tubes had malfunctioned. Malfunctioning pitot tubes were a common occurrence (only in 2008 there was an incident every week concerning pitot tubes), in fact a year prior Airbus had started upgrading these sensors. This makes it a *coherent* candidate. They suggested that supercooled water[15] -a rare phenomenon- froze the pitot tubes, which are under normal conditions equipped to prevent freezing. The freezing of the pitot tubes would have forced the controls (on alternate law) on to the pilot. This led to a precarious situation, so the investigation team argued, given that at high altitudes the speed range that is acceptable is pretty limited, and not an easy task to get right with alarms constantly going off and the pilots having to deal with faulty indications.

The explanation offered by the team is that under these circumstances the pilots did not get the speed right (even 10 knots more or less than the speed they should have gone could lead to a stall). While, as the team showed with simulations, it is possible to maintain control without speed indications of the pitot tubes (by pitching 5° up and setting throttle to 85%), they hypothesized that the pilots failed to do so because they might have missed the low thrust level of the plane with all the alarms going off. This is *coherent* with several past cases where pilots did not correct engine thrust (when auto-thrust was disengaged). One of the possibilities for this, they said, was that in Airbus planes the control stick does not move mechanically (in a Boeing, thrust changes can be felt in the yoke); similarly, they point out that engine power indicators are not situated in the near visual area. This already points to what in the next section will be called *redesign utility*.

As for why the plane did not recover from the stall, crashing into the sea, the team reasoned that the plane might have turned on its side because of the severity of the stall and the turbulence, making the recovery manoeuvre very hard to carry out (especially for untrained pilots).

This 3 stage explanation accounts better for the crash than other alternatives they considered. For example, they consider lighting, but *coherence* rules it out (as stated above), furthermore, it is not *deep* enough, and it is not *complete* (there are error faults sent by the satellite that would lack explanation). They also consider turbulence (rising pockets of air or updrafts as causing the stall). Meteorological maps from NASA show that there were updrafts. But this explanation is not *deep* or *complete* enough (it would not single-handedly explain the stall, though it probably was a contributing factor). Furthermore, they provide their explanation with an eye on redesign, noting improvements such as pilot training, better pitot tubes (which were already being replaced in the Airbus fleet with each new check-up), an improvement in radars (e.g. a mode to detect ice), and the issue of relying too much on automation.

Two years after the crash, the black boxes were found, and the information retrieved partly corrected the explanations given by the team. Rather than attempting to recover from the stall, the pilots never knew that they were stalling. Here is what happened according to the final report of the French civil aviation authority, the BEA:

[14] Traufetter (2010) gave a quite similar account of what transpired during the crash.

[15] Pure water keeps its liquid form at temperatures well below 0C°, but if it touches impurities ice crystals rapidly form.

> The aeroplane went into a sustained stall, signalled by the stall warning and strong buffet. Despite these persistent symptoms, the crew never understood that they were stalling and consequently never applied a recovery manoeuvre. The combination of the ergonomics of the warning design, the conditions in which airline pilots are trained and exposed to stalls during their professional training and the process of recurrent training does not generate the expected behaviour in any acceptable reliable way.(BEA 2012: 200)

A year later, Palmer (2013) gave an even more *complete* and deep diagnosis, listing both the technical causes of the crash, as had the BBC/Nova team, and commenting on the pilots' actions that also contributed to the crash (as listed in the BEA report): first, the pitot tubes froze, causing loss of all speed indications; second, Co-pilot Bonin pulled back his control stick and intermittently maintained it pulled for several minutes, gaining altitude in a nose-up position and thereby stalling; third, the design of the side-sticks made it impossible for other pilots to realize the stick was being pulled, which meant that co-pilot Robert's actions were neutralized with Bonin's; fourth, none of the pilots had received training in high-altitude stall recovery.[16][17]

In sum, what this case illustrates is that at different points in time during investigations of the 2009 Air France Flight 447 crash, different explanatory hypotheses had been advanced, and explanatory virtues – we focused on *simplicity*, *coherence*, *depth*, and *completeness* – played an important role in the weighing and selecting amongst them. This is similar to IBE as analysed in the philosophical literature: the structure of IBE is similar in engineering failure analysis in the sense that the evaluation and selection of explanations hinges on ascribing and weighing of explanatory virtues. In the next section we discuss a salient difference, viz. the predictive information for redesign purposes that engineering malfunction explanations ideally should afford.

5 Redesign utility: An engineering domain-specific virtue in IBE and explanation

There is a salient difference between explanation in engineering science and explanation in other scientific domains: explanation in engineering science is oftentimes a *design driver* for the articulation of improved redesigns. This is a striking difference with other scientific explanations (we are of course not suggesting that explanations in other sciences are not used to further other ends, e.g., policy making, therapeutic and medical interventions, etc.). Engineering explanations are very often not ends in themselves but serve the aim to redesign and modify extant technical systems. In

[16] A contributory factor appears to have been the sophistication of the computerised flight controls. After autopilot disengaged, the pilots were bombarded with an array of information (more than 24 faults in under 4 min, sometimes inconsistent), which did not seem reliable with other measurements (e.g. speed would have read something like 60 knots). As the official report put it, the crew were over-saturated with what seemed like erroneous information. This is not the first incident of this kind – Thomson (2013) showed that there are many cases where a poor interface hinders proper actions.

[17] There were in fact 3 pilots. At the critical moment, the 2 co-pilots (Robert and Bonin) were piloting, while the captain (Dubois) was sleeping. The captain came into the cockpit once it was too late. We know this because he was not wearing a seatbelt (his body was found amongst the debris). The black boxes also confirm this.

engineering failure analysis, this is very prominent. Consider e.g., the redesign (and subsequent manufacturing) efforts that were spent on redesigning the battery ceilings in Samsung Note 7 telephones when explanatory information made clear that the original ceilings were a major cause of the malfunctioning telephones. In such contexts, explanatory information is used to modify extant systems in such a fashion that it is expected that future malfunctions can be prevented. This connection between explanation and redesign is also at play in the Air France case, but in a very tragic way: it was already known – explained – that the pitot tubes were prone to malfunction under extreme weather conditions, and Airbus already had a modified type of pitot tube available which they were already replacing in some units.

In addition to the pitot tubes, the BBC/Nova explanation similarly suggests improvement in radars (e.g. a mode to detect ice). Within this same line of reasoning, Sullenberger (Darlow Smith, Scott 2010) pointed out that there seem to be some serious flaws in the design of the Airbus, such as non-moving auto-throttles, noncentralized indicators (engine power indicators were not situated in the near visual area) and independent controls. Palmer (2013) strengthens this argument by explicitly criticising the design of the side-sticks, that make it impossible for other pilots to realize when the other stick is being pulled. In all these instances, there is either a suggestion for making the design better, so that malfunction would not occur (e.g. radars that detect ice can prevent airplanes flying into storms with said ice), or there is a clear endorsement of an already existing design that would not have the same issues (e.g. Boeing has mechanical yokes that are not independent of each other, which means that if one pilot pulls his control the other will feel it).

On the same note, many have criticized the interface of the cockpit, arguing it does not allow for proper action (the AF447 pilots did not know until the very end that they were stalling). The BEA for instance highlights the responsibility of the ergonomics of the warning design (as well as the conditions in which pilots are trained). Each time Bonin (the pilot) would have pushed the nose down (as he should have), the computer would reactivate (because it would start to read values again), signalling a stall once more, which in such a stressful situation might negatively reinforce that operation. There have been several cases in history where pilots did not know if they were stalling or not based on conflicting information.[18] For this reason, some experts have recommended a display of angle of attack (whose readings are normally only sent to on-board computers), which would help to easily identify and prevent stalling scenarios.

All these examples of redesign suggestions from the case of AF447 show how most engineering explanations are constructed in such a way that the factors that make a difference for the malfunction are contrasted between the current design and an improved design. In other words, most explanations are built with certain redesign aims in mind. This is also why most failure reports have a section on how to improve current design or prevent further occurrences.

This connection between engineering explanation and redesign implies that engineering malfunction explanations also should provide explanatory information of a specific sort that is not required for other scientific explanations. Whereas explanations

[18] Many have pointed out that in alternate law stall protection does not exist, but pilots are often not aware of this, and they probably believed the fly-by-wire system would not allow for such a problem (and that it was a malfunction of the signals).

in other scientific domains should clarify how things depend on one another – e.g., causal dependencies in the case of causal explanations and causal and constitutive dependencies in the case of causal-mechanical explanations – engineering explanations should ideally also provide information that is instrumental in predicting how things relate to one another if certain features of extant technical systems are redesigned and (subsequently) modified. For instance, explanations for the explosion of Samsung Note 7 telephones should clarify how the explosions are dependent on certain features of those technical systems (e.g., battery ceilings of the wrong size) but also provide information that can be used to redesign those systems such that future explosions can be prevented (e.g., battery ceilings with a different size). That is, engineering explanations should offer information that is instrumental in predicting dependencies in to-be redesigned systems – in the example, information about how normal telephone function is dependent on, inter alia, battery ceilings of a certain size. Phrased differently, not only are engineering explanations backward looking by tracking causal dependencies, or synchronic by tracking constitutive dependencies, they should also be forward looking by offering information that enables predicting dependencies that are expected to hold in redesigned systems. So, there is a further engineering domain-specific virtue in terms of which engineering malfunction explanations are evaluated. We call this virtue *redesign utility*: the explanatory information offered should be instrumental in predicting dependencies in redesigned systems.[19] Since predictions are forward looking, the dependencies predicted are *counterfactual* dependencies. This virtue is specific to engineering, marking a remarkable difference with explanation and IBE in other scientific domains.[20]

We further clarify what this virtue entails in terms of a forward-looking counterpart of Woodward's (2003) notion of a what-if-things-had-been-different question, viz. a *what-would-happen-if question* (Weber et al. 2019). In Woodward's (2003) account of causal explanation, explanations should track counterfactual dependencies. For Woodward, adequate causal explanations:

> "locate their explananda within a space of alternative possibilities and show us how which of these alternatives is realized systematically depends on the conditions cited in the explanans. They do this by enabling us to see how, if these initial conditions had been different or had changed in various ways, various of these alternatives would have been realized instead" (Woodward 2003, p. 191).

Woodward unpacks the tracking of counterfactual dependencies between conditions cited in the explanans and alternative explananda possibilities in terms of the idea that

[19] In this paper we focus on explanations for malfunctions that, inter alia, cite faulty designs as causal factors for malfunctioning. Let us stress that the outcome of the explanations is that bad or suboptimal design is a source of the failures. Given such an outcome, we argue, an explanation must be formulated such that it has utility for redesign purposes. It is not the case that the explanatory projects from the outset presuppose faulty designs as causes of the malfunctions. Otherwise, our argument that these explanations should be evaluated in terms of their redesign utility would have a gloss of circularity. That is not the case. We thank a reviewer for urging us to clarify this matter.

[20] Furthermore, redesign is noticeably driven by pragmatic considerations; which is a further difference from the other sciences, who's models generally do not evolve according to pragmatic considerations (such as environmentally friendlier options).

the information provided in explanations should allow us to answer *what-if-things-had-been-different questions*. In the case of causal explanations, you track counterfactual dependencies by looking backwards into the causal ancestry of an event and asking and answering what would have happened to the explanandum phenomenon if its causal ancestry had been different.[21] The greater the range of counterfactual inferences an explanation affords making, the more *depth* it has (Woodward and Hitchcock 2003).[22]

We take it that adequate engineering science explanations ideally do this as well, plus something more: certainly, in the context of malfunction explanation, the explanatory information offered should be instrumental in predicting (forward looking) counterfactual dependencies in redesigned systems. That is, they should offer information that is useful in asking and answering *what-would-happen-if questions* (Weber et al. 2019). We elaborate this claim below. Weber et al. (2019) recently made the case that an important class of technical function ascriptions (what they call "technical advantage function ascriptions") have predictive utility in engineering redesign contexts by giving elaborate answers to questions of the following form:

> What would happen if in *os* [original technical system] component *i* with property *e'* would be replaced by a component *i* that has property *e?*"

For instance, in the case of redesigning the reaction wheel assembly (RWA) of the Hubble space telescope:

> "What would happen if in *os*-RWA bearing balls of bearing assemblies (*i*) that are small (*e'*) are replaced by bearing balls of bearing assemblies that are larger (*e*)*?'*

As Weber et al. (2019) showed with elaborate examples, such questions and their answers provide important information in redesign contexts in which the aim is to redesign extant technical systems that function sub optimally or malfunction, such that redesigned systems will – more precisely, are expected – to function better. Key to such function ascriptions is knowledge about the mechanistic organization of extant systems that are targeted for redesign. For instance, in case of the redesign of the telescope, too much heat was generated in the bearing assemblies of the original system, which was caused by a specific causal chain: frictional forces in the bearing assemblies, which in turn was caused by deformation of the bearing balls, which in turn was caused by the heavy load on the bearings, and this heavy load in turn was caused by the rapid spin of the rotor. Based on this causal-mechanistic information, engineering designers identified the bearing assembly of the RWA as the malfunctioning structure and the size of the bearing balls of the bearing assembly as the feature of this structure responsible for the malfunction: the size of the balls was too small for the load they were intended to

[21] In Woodward's framework, more precisely, the explanatory generalization that figures in the explanans of the explanation.

[22] Woodward and Hitchcock (2003) identify several dimensions of explanatory depth (see also Weslake 2010).

These subtleties need not concern us here.

carry. The proposed redesign solution consisted in a specific increase of the size of the balls. So, what we here see is that causal-mechanistic information on malfunctioning systems or systems that function sub optimally is crucially important in answering what-would-happen-if questions in the context of redesigning such systems. Engineering malfunction explanations provide this information: adequate ones offer information that enables predicting dependencies that are expected to hold in redesigned systems (i.e., they offer information that is instrumental in "technical advantage function ascriptions"). And, arguably, the more *depth* such explanations have, i.e., the greater the range of relevant (forward looking) counterfactual inferences they provide information for, the better they are.[23] Redesign utility is thus an engineering domain-specific virtue in IBE and engineering malfunction explanations.

6 Conclusion

In this paper we charted how inference to the best explanation (IBE) works in engineering science, focussing on the context of malfunction explanation. A first result is that the philosophy of science literature on IBE can be profitably used to understand how IBE works in engineering failure analysis. A second result is that, due to the intimate connection between explanation and redesign in engineering science, there is a salient engineering domain-specific virtue in terms of which engineering malfunction explanations are evaluated, viz. the virtue of *redesign utility*. This virtue entails that the explanatory information offered by a malfunction explanation should be instrumental in predicting counterfactual dependencies in redesigned systems. We illustrated and elaborated these points in terms of a number of engineering examples, focussing in particular on the 2009 crash of Air France Flight 447. In general, our results offer new insights into IBE and explanation by extending analyses of IBE and explanation to engineering science practice. Several relevant follow-up research questions present themselves: how does the virtue of redesign utility operate in different engineering contexts, such as chemical, civil, and mechanical engineering?; how does coherence with other disciplines provide means to borrow explanatory insights from other disciplines profitably to further engineering objectives (e.g. what explanatory and (re)design insights are gained by using laws and equations from physics in engineering explanation and (re)design tasks?). We hope the research presented in this paper offers a direction for thinking about these and related issues.

Acknowledgements We thank Rami Koskinen, Erik Weber, and two anonymous referees for useful feedback on earlier versions of this paper.

[23] Our analysis is complementary to the one of Weber et al. (2019). We identify virtues that malfunction explanations should meet, whereas Weber et al. (2019) assess the utility of predictions in redesign contexts. The source of these predictions can be failure analyses. So, the class of malfunction explanations we consider – those that should meet the virtue of redesign utility – should offer information that is instrumental in predicting dependencies in to-be redesigned systems. These explanations thus, as we argue, should offer information to formulate and answer what-would-happen-if-questions. Weber et al. (2019) assess the structure and utility of these predictive answers.

Funding Funding was provided to Kristian González Barman and Dingmar van Eck by the Research Foundation Flanders (FWO).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval Not applicable.

Informed consent Not applicable.

References

Affonso, L.O.A. (2007). Machinery failure analysis handbook. Gulf Publishing Company.

Amaya, A. (2007). Inference to the best legal explanation. *SSRN Electronic Journal*.

BEA (2012). Final report on the accident on 1st June 2009 to the Airbus A330–203 registered F-GZCP operated by air France flight AF 447 Rio de Janeiro – Paris (pdf), translated by BEA from French, Le Bourget: BEA Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, retrieved 15 May 2019.

Bechtel, W., & Abrahamson, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441.

Bhaumik, S. (2009). A view on the general practice in engineering failure analysis. *Journal of Failure Analysis and Prevention, 9*(3), 185–192.

Bird, A. (2010). Eliminative abduction: Examples from medicine. *Studies in History and Philosophy of Science, 41*, 345–352.

Birolini, A. (2007). *Reliability engineering: Theory and practice* (5th ed.). Berlin: Springer.

Boukharouba, T., Elboujdaini, M., & Pluvinage, G. (Eds.) (2009). *Damage and fracture mechanics. Failure analysis of engineering materials and structures*. Springer.

Craig, J. A., Williams, J.K., Blackburn, G., Linden, S. and Stone R. (2008). Remote detection and real-time alerting for in-cloud turbulence. *AMS 13th Conference on Aviation, Range and Aerospace Meteorology*, 10.5.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press.

Darlow Smith productions, & Scott, K. (2010). *Lost: The mystery of flight 447*. London: BBC.

de Frate, L. (2014). Failure: Analysis of an engineering concept. *Philosophy of Technology.*

Dragulinescu, S. (2016). Inference to the best explanation and mechanisms in medicine. *Theoretical Medicine and Bioethics, 37*(3), 211–232.

Dragulinescu, S. (2017). Inference to the best explanation as a theory for the quality of mechanistic evidence in medicine. *European Journal for Philosophy of Science, 7*(2), 353–372.

Feld, J., & Carper, K. (1997). *Construction failure* (2nd ed.). New York: Wiley.

Fricker, E. (2017). Inference to the best explanation and the receipt of testimony: Testimonial reductionism vindicated. In (Ed.), *Best Explanations: New Essays on Inference to the Best Explanation*. Oxford University press.

Glennan, S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 375–388.

Glennan, S. (2017). The new mechanical philosophy. Oxford University Press.

Hoffmann, M. H. (2010). "Theoric transformations"and a new classification of abductive inferences. *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy, 46*(4), 570–590.

Houkes, W., & Vermaas, P. E. (2010). T*echnical functions: On the use and design of artefacts* (Vol. 1). Springer Science & Business Media.

Illari, P., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*, 119–135.

Jespersen, B., & Carrara, M. (2013). A new logic of technical malfunction. *Studia Logica, 101*(3), 547–581.

Josephson, J. R., & Josephson, G. S. (Eds.). (1994). *Abductive inference*. Cambridge: Cambridge university Press.
Krohs, U. (2009). Functions as based on a concept of general design. *Synthese, 166*, 69–89.
Lipton, P. (2004). *Inference to the best explanation*. Taylor and Francis Group: Routledge.
Loveridge, M. J., Remy, G., Kourra, N., Genieser, R., Barai, A., Lain, M. J., et al. (2018). Looking deeper into the Galaxy (Note 7). *Batteries, 4*(1), 3.
Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 57*, 1–25.
Niiniluoto, I. (1999). Defending abduction. *Philosophy of Science, 66*, S436–S451.
Palmer, W. (2013). Understanding air France 447. Paperback.
Plumbridge, W. J. (2009). New avenues for failure analysis. *Engineering Failure Analysis, 16*(5), 1347–1354.
Preston, B. (1998). Why is a wing like a spoon? A pluralist theory of functions. *Journal of Philosophy, 95*, 215–254.
Psillos, S. (2002). Causation and explanation. *Acumen & McGill-Queens* U.P.
Rapoport, R. (2011). The Rio/Paris crash: Uncovering the secrets that changed aviation history. James Sparling.
Samsung (2017). Samsung Electronics Announces Cause of Galaxy Note7 Incidents in Press Conference [Press release].
Schupbach, J. N. (2017). Inference to the best explanation, cleaned up and made respectable". Best Explanations: New Essays on Inference to the Best Explanation. pp. 39–61.
Schurz, G. (2008). Patterns of abduction. *Synthese, 164*(2), 201–234.
Serrat, O. (2017). The five whys technique. In *Knowledge solutions* (pp. 307–310). Singapore: Springer.
Snooke, N., & Price, C. J. (2012). Automated FMEA based diagnostic symptom generation. *Advanced Engineering Informatics, 26*, 870–888.
Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *Journal of Philosophy, 75*(2), 76–92.
Thomson, J. (2013). *Situation awareness and the human-machine interface*. Safety in Engineering Ltd.
Traufetter, G. (2010). Death in the Atlantic: The last four minutes of air France flight 447. *Spiegel Online*. Retrieved 15 May 2019.
UL. (2017). FAILURE ANALYSIS OF SAMSUNG NOTE 7 [Press release].
van Eck, D. (2015). Mechanistic explanation in engineering science. *European Journal for Philosophy of Science, 5*(3), 349–375.
van Eck, D. (2016). The philosophy of science and engineering design. Springer.
van Eck, D., & Weber, E. (2014). Function ascription and explanation: Elaborating an explanatory utility desideratum for ascriptions of technical functions. *Erkenntnis., 79*(6), 1367–1389.
Van Eck, D., & Weber, E. (2017). In defense of coexisting engineering meanings of function. *AI EDAM, 31*(1), 55–68.
Weber, E., van Eck, D., & Mennes, J. (2019). On the structure and epistemic value of function ascriptions in biology and engineering sciences. *Foundations of Science, 24*(3), 559–581.
Weslake, B. (2010). Explanatory depth. *Philosophy of Science, 77*(2), 273–294.
Williams, J. K., Sharman, R. and Kessinger, C. (2009). Developing a global turbulence and convection Nowcast and forecast system. Published by the American Institute of Aeronautics and Astronautics.
Woodward, J. (2003). Making things happen: A theory of causal explanation. Oxford University Press.
Woodward, J., & Hitchcock, C. (2003). Explanatory generalizations, part I: A counterfactual account. *Noûs, 37*(1), 1–24.
Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies, 148*(2), 201–219.
Yun, J. J., et al. (2018). Benefits and Costs of Closed Innovation Strategy: Analysis of Samsung's Galaxy Note 7 Explosion and Withdrawal Scandal. *Journal of Open Innovation: Technology, Market, and Complexity, 4*(3), 20.