Forthcoming in V. C. Müller (ed.), Fundamental Issues of Artificial Intelligence (Synthese Library), Berlin: Springer. Penultimate Draft.

Ad Hoc Hypotheses and the Monsters within

Ioannis Votsis
New College of the Humanities / Heinrich-Heine University Duesseldorf
ioannis.votsis@nchum.org / votsis@phil.hhu.de

Abstract

Science is increasingly becoming automated. Tasks yet to be fully automated include the conjecturing, modifying, extending and testing of hypotheses. At present scientists have an array of methods to help them carry out those tasks. These range from the well-articulated, formal and unexceptional rules to the semi-articulated and variously understood rules-of-thumb and intuitive hunches. If we are to hand over at least some of the aforementioned tasks to machines, we need to clarify, refine and make formal, not to mention computable, even the more obscure of the methods scientists currently employ in their inquiries. The focus of this essay is one such less-than-transparent methodological rule. I am here referring to the rule that ad hoc hypotheses ought to be spurned. This essay begins with a brief examination of some notable conceptions of ad hoc-ness in the philosophical literature. It is pointed out that there is a general problem afflicting most such conceptions, namely the intuitive judgments that are supposed to motivate them are not always shared. Instead of getting bogged down in what ad hoc-ness exactly means, I shift the focus of the analysis to one undesirable feature often present in alleged cases of ad hoc-ness. I call this feature the 'monstrousness' of a hypothesis. A fully articulated formal account of this feature is presented by specifying what it is about the internal constitution of a hypothesis that makes it monstrous. Using this account a monstrousness measure is then proposed and compared with the minimum description length approach.

1. Introduction

Science is increasingly becoming automated. Tasks yet to be fully automated include the conjecturing, modifying, extending and testing of hypotheses. At present scientists have an array of methods to help them carry out those tasks. These range from the well-articulated, formal and unexceptional rules to the semi-articulated and variously understood rules-of-thumb and intuitive hunches. If we are to hand over at least some of the aforementioned tasks to machines, we need to clarify, refine and make formal, not to mention computable, even the more obscure of the methods scientists currently employ with some measure of success in their inquiries. The focus of this essay is one such less-than-transparent methodological maxim for which much confusion and disagreement persists. I am here referring to the maxim that ad hoc hypotheses ought to be spurned. The need to throw light on this maxim and, in particular, on the notion of ad hoc-ness becomes all the more obvious when one considers that it is routinely invoked by scientists in assessing modifications or extensions of existing hypotheses and conjectures of new ones.

This essay begins with a brief examination of some notable contributions to the philosophical literature on ad hoc-ness, focusing, in particular, on two prominent conceptions. After raising some problems specific to these conceptions, it is pointed out that there is a more general problem afflicting most conceptions in the philosophical literature, namely the intuitive judgments that are supposed to motivate such conceptions are not always shared and indeed are sometimes even in conflict. The concept of ad hoc-ness is already burdened with too much intuitive baggage, signifying too many different things to too many different people. Instead of getting bogged down in what ad hoc-ness exactly means, I shift the focus of the analysis to one undesirable feature often, but not universally, present in alleged cases of ad hoc-ness. I call this feature the 'monstrousness' of a hypothesis for reasons that will become

clear below. A fully articulated formal account of this feature is given by specifying what it is about the internal constitution of the content of a hypothesis that makes it monstrous. Using this account a monstrousness measure is then proposed and compared with the minimum description length approach to simplicity/ad hoc-ness that is so popular in information and computer learning theory. It is argued that the monstrousness measure has some definite advantages over the minimum description length approach, at least in so far as the latter is traditionally conceived, but also some disadvantages. The main such disadvantage is that, as it stands, the measure is impracticable. The essay concludes with a proposal that seeks to extract a more practicable version of the measure. In doing so, the hope is that this essay helps prepare the ground for the delegation of a full gamut of scientific duties to the machines of the future.

2. Ad hoc-ness

There is a hefty amount of confusion surrounding the notion of ad hoc-ness. Given the notion's prevalence in everyday discourse, a dictionary entry makes for an apt starting point to our investigation. According to the Oxford English Dictionary, 'ad hoc' means "formed, arranged, or done for a particular purpose only".[1] This ordinary conception of ad hoc-ness is reflected in compound expressions like 'ad hoc committee'. But what does it mean for a hypothesis to be ad hoc? By and large, philosophers of science and other interested parties to this debate find the ordinary conception of ad hoc-ness deficient as an answer to this question. That's where their agreement ends, however, as they quarrel over how best to go about answering it. As a result, several conceptions of ad hoc-ness have arisen through the years.

Although the philosophical literature on ad hoc-ness is far from vast, there is more than enough material to prevent one from doing it justice in a short essay like this. I will thus restrict my comments to some notable contributions. Aside from Popper (1972) and Zahar (1973), both of which will be discussed in a bit more detail below, I would like to mention three other contributions in passing. The first is Leplin (1975). He identifies no less than five individually necessary and jointly sufficient conditions for ad hoc-ness, namely experimental anomaly, justification, tentativeness, consistency and non-fundamentality. Alas, his view suffers also from a number of serious problems (see, for example, Grünbaum 1976) and is ultimately premised on the misguided idea that there is a unique "concept of '*ad hoc*ness' which the scientific community employs" (p. 316) [original emphasis]. Why it is misguided will, I hope, become clear in the next section. The second is Grünbaum (1976). He considers three temporally-relativised notions of ad hoc-ness, which he calls ad hoc (a), (b) and (c). To give the reader a taste, the first one is meant to cover those cases where a hypothesis is ad hoc at the moment of its introduction. This means, roughly, that it has neither theoretical nor independent empirical support or disapprobation at that moment. The second and third notions are progressively stronger, the latter being so strong that its historical instantiation is in doubt. Ultimately, none of this matters as Grünbaum argues, for the most part quite convincingly, that attempts to de-temporalise these notions and hence to articulate them in purely logical terms are likely to fail. The upshot is that Grünbaum doubts the view that there exists a precise and adequate conception of ad hoc-ness. The third and final contribution is that of Forster and Sober (1994). They contend that the term 'ad hoc' signifies unreasonable revisions to scientific theories and, in particular, to auxiliary hypotheses. Though they do not attempt to pin down the notion, they allude to an important connection between non-ad hoc-

---

[1] The same dictionary traces the Latin expression, which literally means 'for this', to the middle of the sixteenth century. See: http://www.oxforddictionaries.com/us/definition/american_english/ad-hoc?q=ad+hoc

ness and simplicity.[2] The suggestion, roughly, is that simpler hypotheses are less ad hoc. We return to this connection in Section 5 below.

Let us now turn to Popper (1972). Popper's conception attempts to unpack the specificity of an ad hoc hypothesis in terms of its lack of excess testable content. In his own words:

> *Ad hoc* explanations are explanations which are not independently testable... In order that the *explicans should not be ad hoc, it must be rich in content: it must* have a variety of testable consequences, and among them, especially, testable consequences which are different from the *explicandum. It is these different* testable consequences which I have in mind when I speak of *independent tests, or* of *independent evidence* (1972, pp. 15-16, 193) [original emphasis].

Its connection to the ordinary conception of ad hoc-ness should be obvious. If a hypothesis *H* has no excess testable content over explicandum *E*, then its purpose seems at best restricted to that of attempting to explain *E*. If, however, it has excess testable content, then the hypothesis has a broader purpose in that it can potentially explain other evidence that may turn out to be true. Following Popper, we may illustrate this conception with an example from the history of science. The hypothesis positing the existence as well as the orbital and mass characteristics of the planet Neptune helps explain the perturbed orbit of Uranus within the Newtonian paradigm. That same hypothesis, however, has excess testable content over and above the perturbed orbit of Uranus. Among other things, it predicts (indeed to some extent successfully) additional perturbations in the orbits of other planets as well as that of the Sun around the solar system's barycenter.

Various problems afflict Popper's conception, one of which will be mentioned here. The problem I have in mind questions the idea that excess testable content is a sufficient condition for non-ad hoc-ness. Take *any* explicans that we would all, or at least the Popperians, judge to be ad hoc. We can easily turn it into one that Popperians would deem non-ad hoc simply by conjoining to it *any* random proposition whose testable content exceeds that of the explicandum. Here's an example. Suppose that the following hypothesis Z_1 offers an ad hoc explanation of S_1.

Z_1: Zeus exists and he is sometimes angry and whenever he is angry he lights up the sky with thunderbolts.

S_1: Sometimes the sky lights up with thunderbolts.

Suppose, moreover, that we add a random proposition A_1 to the explicans. It doesn't matter whether this proposition is true or false – in this case we happen to choose a true one.

A_1: Free falling objects near the earth's surface accelerate roughly at 9.81 m/s^2.

Following Popper's conception, conjunction $Z_1 \wedge A_1$ is not an ad hoc explanation of S_1 for it has excess testable (and in fact tested) content, namely that concerning the rate of acceleration of freely falling objects near the surface of the earth. In other words, having

---

[2] Simplicity, according to Forster and Sober, ought to be understood formally in terms of the Akaike information criterion. This criterion provides a method for selecting hypotheses by estimating their predictive accuracy. It does so by taking into account the trade-off between the simplicity of a hypothesis and its goodness-of-fit toward the data. For more details, see Akaike (1974).

excess testable content does not guarantee non-ad hoc-ness. At best, having excess testable content is a necessary condition for non-ad hoc-ness.[3]

Consider next Zahar's approach to ad hoc-ness. I here focus on one of three proposed conceptions put forth by him.[4] A theory, holds Zahar, is ad hoc "if it has no novel consequences as compared with its predecessor" (1973, p. 101). A consequence or prediction is novel so long as the corresponding phenomenon was not the explanatory target of the scientists who constructed the theory. For example, the perihelion of the planet Mercury is a novel prediction of the general theory of relativity because the perihelion was presumably not explanatorily targeted by Einstein in his construction of the theory.[5] There are some differences between Zahar's and Popper's conceptions. For example, Zahar construes ad hoc-ness as a relation between successive hypotheses, not, as Popper does, simply as a kind of failed explanatory relation. Even so, ad hoc-ness judgments elicited from the two conceptions are, on the face of it at least, often in agreement. A hypothesis possessing consequences whose corresponding phenomena were not explanatorily targeted during construction has excess testable content over the phenomena it was constructed to explain. That's not to say that judgments elicited from the two conceptions are never divergent however. A successor hypothesis with a solitary novel consequence has excess testable content and hence is non-ad hoc for Popper but counts as ad hoc for Zahar if that consequence is also novel for its predecessor.[6]

Zahar's conception fares no better than Popper's. Two problems stand out. First, just because a successor hypothesis makes no novel predictions compared to its predecessor does not mean that it is any worse off or indeed ad hoc. Some progress in science involves removing ad hoc elements from a predecessor hypothesis to obtain a non-ad hoc, or less as hoc, successor hypothesis. Ridding an otherwise empirical hypothesis from supernatural posits is exactly one such type of progress. Using the Zeus hypothesis as an example, this would mean shifting from Z_1 (the predecessor hypothesis) to the (admittedly meagre) S_1 (successor hypothesis). Second, suppose that a predecessor hypothesis H explains all and only O_1. Suppose moreover that two scientists, F and G, independently construct the same successor hypothesis H', which explains all and only O_1 and O_2. Suppose, finally, that O_2 was explanatorily targeted by G but not by F. Eliciting a judgment from Zahar's conception of ad hoc-ness yields a contradiction: H' turns out to be both ad hoc and non-ad hoc.[7]

---

[3] Grünbaum (1976, p. 343) notes Popper's ambiguous behaviour towards the logical status of the condition of excess testable content, sometimes treating it as merely a sufficient condition and sometimes as both necessary and sufficient for non-ad hoc-ness.

[4] Here are the other two: "It is *ad hoc₂* if none of its novel predictions have been actually 'verified'... [a] theory is said to be *ad hoc₃* if it is obtained from its predecessor through a modification of the auxiliary hypotheses which does not accord with the spirit of the heuristic of the programme" (1973, p. 101) [original emphasis]. Grünbaum (1976, p. 341) notes the similarities between Zahar's notion *ad hoc₁* and his own notion *ad hoc* (c). For a more detailed discussion of Zahar's notions, particularly *ad hoc₂*, the reader may consult Redhead (1978).

[5] This example has been contested by Earman and Glymour (1978) who argue, convincingly, that the perihelion was in fact explanatorily targeted. Less controversial examples include all the cases involving temporally novel phenomena, i.e. phenomena which were not known at the time a hypothesis was constructed and hence could not have been explanatorily targeted by the constructors.

[6] How much divergence exists between judgments elicited from the two conceptions depends on a number of factors. For example, if Zahar permits the notion of novel consequences to also range over consequences that cannot be tested, then the divergence is significant.

[7] Strictly speaking, H' is ad hoc for G but non-ad hoc for F. If Zahar were a subjectivist then he could perhaps get away with this reply by claiming that ad hoc-ness is a subjective matter. The trouble is he is not – see his comments (1973, pp. 103-104). What is more, if ad hoc-ness is to have any epistemic import it could not be something that varies from subject to subject. For more on this and related problems see Votsis (2014).

3. Taking a Step Back

Is it not immature to stop our investigation into conceptions of ad hoc-ness after considering only two of them? Yes, but there is good reason for doing so. The existing conceptions, as well as at least some of the objections levelled against them, rely heavily on intuitive judgments about what are genuinely ad hoc or non-ad hoc cases and what are essential and what accidental features of ad hoc-ness. The problem is that these intuitive judgments are not always shared. In fact, some such judgements are sometimes in conflict. As an illustration of this conflict, take the intuition that ad hoc-ness is always undesirable. First off, note that this intuition is not inherited from the ordinary conception of ad hoc-ness. Ad hoc committees, for example, serve a more limited role than non-ad hoc ones but that doesn't make them any less undesirable. Secondly, and more importantly, it is highly doubtful that there is a uniform meaning of ad hoc-ness in the scientific context. Forster and Sober, for instance, are categorical in their condemnation of ad hoc hypotheses: "we reserve the term 'ad hoc' for revisions of the bad kind" (1994, p. 17). But not everybody agrees. Grünbaum, for example, holds that "I evidently do not deny that *certain* uses of the term 'ad hoc' are intended to be derogatory" (1976, p. 361) [first emphasis added]. And Popper is of two minds. He sometimes claims that "the corresponding requirement that explanations of this kind [i.e. ad hoc] should be avoided are, I believe, among the main motive forces of the development of science" (1972, p. 192). But at other times he suggests that "... we must not exclude all immunizations, not even all which introduce *ad hoc* auxiliary hypotheses" (1974, p. 32). More generally, and perhaps more interestingly, Holton (1969) puts paid to the idea that scientists themselves share a unique conception of what it means for a hypothesis to be ad hoc:

> The scientist who adopts somebody's hypothesis or creates his own for a specific purpose, "in order to account" for a bothersome result or feature of the theory, regards it as ad hoc - not necessarily in a derogatory sense... Thus we have found in the scientific literature characterizations of the following kinds for acceptable ad hoc hypotheses: "not inconceivable," "reasonable," "plausible," "fundamental," "natural," "appealing," "elegant," "likely," "assumed a priori to get the desired results," "auxiliary" or "working hypothesis." On the other hand, when an ad hoc hypothesis is rejected, we see it described in the following way: "artificial," "complex," "contrived," "implausible," "bothersome," "unreasonable," "improbable," "unlikely," "unnecessary," "ugly." (p. 178).

How do we proceed in light of such discord? Instead of attempting to disentangle the intuitions behind the use of the term 'ad hoc-ness', I propose that we focus on one feature – or, otherwise put, one consistent set of intuitions – that is often associated with alleged cases of ad hoc-ness.[8] This is a feature that I deem to be undesirable at all times for a hypothesis to possess. For this reason it is useful that it be given a clear conception. After all, in matters of sound methodology and epistemology we need concepts and rules that tell us in as unambiguous a way as possible what is a legitimate and what an illegitimate modification of a central hypothesis and/or an auxiliary, which hypotheses are likely to be true and which false, what evidence weighs more and what less, etc. In order to avoid any carry-over from the intuitive baggage associated with the term 'ad hoc', or the intuitive baggage of any other closely related term for that matter, I suggest that we use a relatively unsullied term to express this feature. I call it 'monstrousness' for reasons that will soon become apparent.[9]

---

[8] One possible approach which is eschewed here is to admit several distinct notions of ad hoc-ness. In my view this is a legitimate approach but one that does not focus on

[9] A version of this notion was first explored in Votsis (2014).

4. Disjointedness and Monstrousness

The characteristic I have in mind is something that has not gone unnoticed in discussions of ad hoc-ness. Grünbaum, for example, asserts that what is undesirable about some ad hoc hypotheses is that they are "artificial, contrived or arbitrary" (1976, p. 358). Instead of trying to expand on what any of these terms truly mean, a move which will inevitably pull us back down into controversy, I will instead employ the largely untainted term 'monstrousness'. The reason why we call this characteristic, the 'monstrousness' of a hypothesis, is that it indicates the extent to which a hypothesis is assembled out of confirmationally disjointed content parts, in a manner similar to the way the most famous monster in literature, i.e. the monster in Mary Shelley's Frankenstein, is assembled out of a motley of parts.[10] Numerous examples of undesirable ad hoc hypotheses exhibit this disjointedness in good measure. This is certainly true of Z_1. It is, however, less pervasive and perhaps even absent in hypotheses that are typically more desirable, e.g. the conjunction of the Newtonian paradigm with the hypothesis that Neptune exists and has certain orbital and other characteristics.

How are we meant to understand disjointedness? I propose the following articulation:

Disjointedness: Any two content parts expressed as consistent propositions *A*, *B* are disjointed if and only if (i) $P(\alpha/\beta) = P(\alpha)$ for all propositions α, β where α is a relevant (and non-redundancy containing) deductive consequence of *A* and β is a relevant (and non-redundancy containing) deductive consequence of *B* and (ii) for any such pair α, β, there is no proposition γ that is a relevant deductive element of $\alpha \wedge \beta$.

Hence jointedness can be articulated thus:

Jointedness: Any two content parts expressed as consistent propositions *A*, *B* are jointed if and only if (1) $P(\alpha/\beta) \neq P(\alpha)$ for some pair of propositions α, β where α is a relevant (and non-redundancy containing) deductive consequence of *A* and β is a relevant (and non-redundancy containing) deductive consequence of *B* or (2) for some such pair α, β, there is a proposition γ that is a relevant deductive element of $\alpha \wedge \beta$.

Let us take a closer look at the notion of disjointedness. Consider clause (i) first. The probabilities involved are meant to be objective. That is, they are meant to indicate true relative frequencies and/or true propensities of events, states-of-affairs, properties, etc., expressed by propositions. This is an important qualification as it de-subjectivises the notion of disjointedness (and hence jointedness) – we briefly return to this issue in the next section. The notion of probabilistic independence allows us to make an important first step in expressing the idea that two content parts are confirmationally disjointed. For if α, β are probabilistically independent we know that the degree of truth of the one, i.e. its probability, is not affected if we assume something about the truth (/falsity) of the other.[11] This connects well with the relevance criterion of confirmation various Bayesians live by according to which *e* stands in a confirmational relation to (i.e. either confirms or disconfirms) *h* if and only if the two are probabilistically dependent.

---

[10] For a discussion of the notion of content part you may consult Gemes (1994; 1997).

[11] To avoid the usual problems, and following common practice, false propositions must be assigned a non-zero value.

Properly accounting for the confirmational disjointedness of two propositions A, B requires that we inspect not only their total content but also the content of their parts. That's because the two propositions may be probabilistically independent even though some of their parts are not. That's where the notion of *deductive consequence* comes in handy. By checking whether each and every – see the next paragraph for a qualification – deductive consequence of the one proposition is probabilistically independent from/dependent on each and every deductive consequence of the other we ensure that we take all confirmation relations between A, B into account. Consider the following example as an illustration. Suppose that A: $U_1 \wedge U_2$ and B: $U_1 \wedge V_1$, that $P(U_1/U_2)=P(U_1)$ and $P(U_1/V_1)=P(U_1)$ and that $P(U_1) = 0.5$, $P(U_2) = 0.5$, $P(V_1) = 0.5$ and $P(B/A) = 0.25$. From the fact that U_1 and U_2 are probabilistically independent we can derive that $P(U_1 \wedge U_2) = P(U_1) * P(U_2) = 0.25$. So we know that $P(A) = 0.25$. Similarly from the fact that U_1 and V_1 are probabilistically independent we can derive that $P(U_1 \wedge V_1) = P(U_1) * P(V_1) = 0.25$ and thus we know that $P(B) = 0.25$. Using Bayes theorem we can thus derive that $P(A/B) = 0.25$. But that just means that A, B are probabilistically independent since $P(A/B) = P(A)$. But notice that the same is not true of all the consequences of A, B. Take U_1. It is a consequence of A and of B. But $P(U_1/U_1) = 1$ and hence $P(U_1/U_1) \neq P(U_1)$. Thus, there is a proposition α and a proposition β, namely U_1 in both cases, such that $P(\alpha/\beta) \neq P(\alpha)$ and hence A, B turn out to be jointed when we take a closer look. To recap, in order to make sure that two propositions are confirmationally disjointed we must demand that probabilistic independence holds *all the way down*.

As already alluded, we do not actually care about *all* deductive consequences. This is because some of them are trivial. In fact, were we to take these into account we would render the concept of disjointedness unsatisfiable. This can be demonstrated with a simple example. Regardless of the exact content of A, B there are always validly derivable but trivial consequences of each that they have in common, e.g. $A \vee B$. Such trivial common consequences guarantee the existence of a pair of propositions α_i, β_i for which $P(\alpha_i/\beta_i) \neq P(\alpha_i)$ provided $0 < P(\alpha_i) < 1$. Otherwise put, it guarantees that A, B are not disjointed. To rule out such cases we restrict our attention to all *relevant* deductive consequences. The notion of relevance can be found in Schurz (1991) where he explains that "the conclusion of a given deduction is irrelevant iff the conclusion contains a component [i.e. a formula] which may be replaced by any other formula, salva validitate of the deduction" (pp. 400-401).[12]

Note that although two propositions A, B may be probabilistically independent all the way down, they may still be confirmationally related through jointly and relevantly – as in 'having a relevant deductive consequence' – entailing a proposition γ that neither entails on its own and whose truth would confirm both.[13] Clause (ii) is there to ensure that there is no 'indirect' confirmational relation between A and B via such a consequence γ. If there is such a consequence then A, B are not disjointed. Not just any joint consequence will do. Unless we place some restrictions on what counts as a proposition γ, the concept of disjointedness is once again rendered unsatisfiable. That is, if we required only that there is no γ that any α, β

---

[12] There is also an analogous notion that applies to predicates – see Schurz (2014).

[13] An example can be garnered from discussions of causal modelling. Two causes may be probabilistically independent and yet their presence may be sufficient to yield a joint effect. In many cases the presence of the joint effect confirms the presence of both causes. Thus, the two causes are confirmationally related even though they and the propositions expressing their presence are probabilistically independent – the latter presumably all the way down. Another example may be sourced from the domain of mathematics. Axioms are probabilistically independent (presumably all the way down) but two or more of them may be necessary to derive a single theorem. This last example is only meant as a crutch to help understand condition (ii). The view I am proposing here is restricted to empirical, not mathematical, hypotheses.

jointly (but not individually) entail then there would always be such a γ, no matter what content A, B possess. For example, one such γ is $\alpha \wedge \beta$. More generally, a proposition ζ that is jointly (but not individually) entailed by two propositions α, β cannot play the role of γ if it is logically equivalent to $\delta \wedge \varepsilon$ where either δ is a relevant consequence of α or ε is a relevant consequence of β or both. Such joint consequences are trivial for our purposes and therefore incapable of assisting is us in our quest to find a confirmational relatedness between A, B that is not captured by condition (i). Note that to dismiss such trivial consequences we cannot rely on the simple notion of relevant consequence like we did before since $A \wedge B$ is a relevant deductive consequence of $A \wedge B$, i.e. it is not the case that we can substitute any formula in it with any other formula without disturbing the validity of the deduction. Instead, to solve our problem we must rely on the associated notion of a relevant (deductive) element. A relevant (deductive) element of a proposition φ is a relevant consequence of φ that is also an element of φ. What is an element of φ? Roughly speaking, a content element is a proposition whose content is so small that it cannot be decomposed into smaller content parts.[14] Such a proposition could not, for example, be logically equivalent to a conjunction of two probabilistically independent propositions since that would mean that the content of each is smaller than the original proposition. By requiring γ to be a content element we in effect rule out trivially derivable joint consequences like $\alpha \wedge \beta$ and indeed all the aforementioned $\delta \wedge \varepsilon$ pairs since they can always be decomposed into propositions with content parts that are smaller than γ, namely δ and ε.

Finally, consider the non-redundancy qualification. This is there to simplify things. By redundancy-containing propositions, I mean propositions that repeat content. e.g. $\alpha \wedge \alpha$. Note that if some α_i is a relevant deductive consequence of A, then $\alpha_i^1 \wedge ... \wedge \alpha_i^n$ where $n > 1$ is also a relevant deductive consequence of A, where the subscript i denotes a given proposition and the superscripts denote different syntactic instances of that proposition. Such redundancy-containing propositions as $\alpha_i^1 \wedge \alpha_i^2$ need not be considered in our evaluations of probabilistic (in)dependence. This is because $P(\alpha^1 \wedge ... \wedge \alpha^n) = P(\alpha^1)$ for any proposition α and hence $P(\alpha_i/\beta_i) = P(\alpha_i)$ for some pair α_i, β_i if and only if $P(\alpha_i^1 \wedge ... \wedge \alpha_i^n / \beta_i^1 \wedge ... \wedge \beta_i^m) = P(\alpha_i^1 \wedge ... \wedge \alpha_i^n)$ for any n, $m > 1$, again where the superscripts denote different syntactic instances of the same proposition. The non-redundancy clause thus ensures we waste no time comparing redundancy-containing relevant deductive consequences.

What use could we possible have for disjointedness? Well, disjointedness forms a barrier against the spread of confirmation. Thus even though monstrous hypotheses get confirmed under this view, the confirmation they receive for a content part that is disjointed from other content parts doesn't spread to those other parts. For example, the truth of S_1 confirms (a part of) Z_1 precisely because Z_1 was designed to entail S_1 but, crucially, this confirmation does not spread to the non-S_1 part of Z_1, namely the part that asserts that Zeus exists and posits the existence of Zeus and the property that he is sometimes angry. The approach just outlined is similar to Schurz (2014) in that it aims to regulate how confirmation spreads within the content parts of a hypothesis. Unlike him, however, I insist that we are still dealing with a case of *genuine* confirmation when the support gained from some piece of evidence does not spread to content parts other than that corresponding to the evidence. My approach thus offers a unified treatment of confirmation relations.

---

[14] For a detailed exposition, see Schurz and Weingartner (2010).

Hypotheses may possess both disjointed and non-disjointed content parts. To be exact, since disjointedness and non-disjointedness are relations that hold between various content parts of hypotheses, the claim is that hypotheses may possess content parts, some of which are disjointed and others non-disjointed to other content parts. It is my conjecture that almost all hypotheses will have some disjointed parts and hence will be monstrous to some extent. It thus makes sense to devise a way to measure the level of monstrousness of a hypothesis. But before we do that we must consider one last complication.

Should judgments of monstrousness be affected by the way the same content is distributed between two propositions? We are obviously free to cut up content any which way we like, i.e. content distribution is an arbitrary affair. Being arbitrary means that it doesn't tell us anything about the world. Hence, in answer to the above question, the way in which content is distributed should not affect our monstrousness judgements. Take two propositions A_1, B_1 where A_1: $D_1 \wedge D_2 \wedge D_3$ and B_1: E_1. Suppose that when we compare these propositions using some measure of monstrousness we get some score w_1. Now take another two propositions A_1', B_1' where A_1': $D_1 \wedge D_2$ and B_1': $E_1 \wedge D_3$. Suppose that when we compare A_1', B_1' using the same measure of monstrousness the resulting score is w_2. According to the above argument, since $A_1 \wedge B_1$ has the same content as A_1' $\wedge$ B_1' any proposed measure of monstrousness should ensure that $w_1 = w_2$. That is to say, monstrousness judgments should be invariant under different ways of distributing the same content between two propositions.

One, perhaps the only, way to pull this off is to calculate monstrousness scores on the basis of *all* distinct ways of distributing the same content between two propositions. In what follows, I put forth a proposal of exactly such a measure. The proposed measure is not the final word on the matter but still worth considering since, in my view, it is heading in the right direction.[15] Without further ado, here's the proposal: The monstrousness of two propositions is given by the ratio of the sum of disjointed pairs of parts taken from all distinct ways of distributing content to the sum of the total number of pairs of parts (i.e. jointed and disjointed) taken from all distinct ways of distributing content. Formally, the monstrousness m of two propositions A, B is given by the following function:

$$m(A, B) = \sum_{i=1}^{n} d_i^{A, B} \Big/ \sum_{i=1}^{n} t_i^{A, B}$$

where $d_i^{A,B}$ denotes the number of disjointed pairs in a given content distribution i, $t_i^{A,B}$ denotes the total number of jointed plus disjointed pairs in a given distribution i, the superscript A, B denotes that each i is a distribution of the same content as that contained in $A \wedge B$ and n denotes the total number of content distributions.[16]

The number of disjointed pairs in a given content distribution is determined by counting how many times a different pair of relevant deductive consequences α, β turn out to be probabilistically independent *and* for which no proposition γ that is a relevant deductive element of $\alpha \wedge \beta$ exists. Any pair that is not disjointed is counted as jointed. The higher

---

[15] One reason for its inadequacy is that when a pair of propositions α, β are deemed jointed the strength of their jointedness, e.g. the degrees of their probabilistic dependence, is neglected. Such information should surely have a role in any suitably sensitive measure of monstrousness.

[16] This function only makes sense if the total number of content distributions is finite. I am assuming this is the case. Arguments for this assumption can be given but require quite a bit of stage-setting. For now, it suffices to say that this assumption is guaranteed to hold if the hypotheses in question can be fully decomposed into a finite number of content elements.

(/lower) the value of *m* the more (/less) monstrous the content expressed by $A \wedge B$. Note that this value is the same no matter how we cut *A, B* since we take into account *all* other ways the same content can be distributed between two propositions.

One of the advantages of the proposed measure is that it is quite broad in its range of application. This is due to the fact that the notions of jointedness and disjointedness don't place any restrictions on the propositions being compared other than the restriction that they are consistent. As a result, the said propositions can be drawn from a large pool of entries which includes central hypotheses, auxiliaries, explanantia and explananda. This not only allows us to gauge the monstrousness of the most commonly touted relations, e.g. the relation between a central hypothesis and an auxiliary hypothesis or the relation between an explanans and an explanandum, but also of any other relation we can think of, e.g. the relation between one auxiliary hypothesis and another. But what about individual hypotheses? How does a *comparative* measure like monstrousness pass judgment on *individual* hypotheses? Without difficulty as it so happens. So long as such hypotheses are not content elements, we can split their content into two propositions which can then be compared.[17] Thus even individual hypotheses have a monstrousness score. This should not be surprising as whether or not a given proposition forms part of, say, a central hypothesis or is a detached auxiliary seems to be (at least occasionally) a matter of convention.

5. Monstrousness and MDL

This essay is meant to give some guidance, however limited, on the subject of how machines may automate the task of discriminating between bad and good hypotheses. With this aim in mind it is worth comparing, albeit briefly and superficially, my approach to a leading approach employed in information and computer learning theory, namely the minimum description length (MDL) approach. MDL is a hypothesis selection principle. The best hypothesis, according to this approach, is the one that provides the most economical compression of the data. It is not hard to see how MDL is related to the demand for simpler hypotheses. Thus, Rissanen, MDL's founding father, doesn't hesitate to assert that "the notion of simplicity is entirely in line with the modern notions of complexity of description" (1983, p. 421). The notion of ad hoc-ness is not as prominent in this literature. But when it does make its appearance the central idea seems to be that simpler hypotheses are less ad hoc – an idea that, as we earlier saw, Forster and Sober also find attractive.

> ... the *minimum description length principle...* extends Occam's Razor to say that the best hypothesis is the one that minimizes the total length of the hypothesis plus the description of the *exceptions* to the hypothesis. The intuition is that ad hoc hypotheses are simply lists of examples, which makes them no shorter than the examples they purport to summarise. In contrast, good hypotheses reduce many examples to a simple, general rule (Shavlik and Dietterich 1990, p. 47) [original emphasis].

There are obvious similarities between my measure of monstrousness and MDL measures of simplicity. For example, a hypothesis that conjoins propositions that express disparate facts, e.g. a_1 is a white swan $\wedge$ a_2 is a white dwarf, gets a high monstrousness score *and* a low MDL simplicity score since its length is presumably no shorter than the combined length of the

---

[17] Two points are worth making. First, as we saw earlier any such division is adequate as all divisions must be considered before a monstrousness score is produced. Second, our exclusion of hypotheses that are content elements is inconsequential since such hypotheses cannot but be wholly non-monstrous given that they have no distinct content parts that can be disjointed.

individual propositions it seeks to summarise.[18] But there are also differences. A hypothesis that conjoins propositions that express related facts, e.g. b_1 is a white swan $\wedge$ b_2 is a white swan $\wedge$ ... $\wedge$ b_n is a white swan, but does not convey them in terms of a generalisation, e.g. all observed swans are white, gets a low simplicity score for the abovementioned reason but *does not* get a high monstrousness score. The reason for the latter is that the individual facts are not probabilistically independent all the way down since they are systematically related via common relevant deductive consequences, e.g. the claim that there is a white swan. A corollary is that under the monstrousness approach, and contra MDL, a conjunction of propositions expressing related facts gets the same score as a generalisation of them since they both possess the same content and hence share the same distinct ways of distributing that content between two propositions.

Two further differences between these approaches, this time at a more abstract level, are worth pondering over. Monstrousness is determined by objective facts about the true relative frequencies and/or true propensities of events, states-of-affairs, properties, etc., expressed by the propositions being compared. Its determination is thus a thoroughly *a posteriori* matter and herein lies the strength of this approach. By contrast, the MDL approach to simplicity seems to reward or penalise hypotheses in a strongly *a priori* manner by insisting that the simplest hypothesis is the best regardless of facts on the ground. That's the first difference between the two approaches. The second difference concerns practicality. Here the earlier mentioned strength of the monstrousness approach turns into a weakness. Since the probabilities involved are objective and we have limited access to these the approach is severely handicapped in its practicability. The same is not true of the MDL approach, which, by virtue of its strong *a prioricity*, is capable of offering advice even in the face of limited access to facts on the ground.

Allow me to draw this section to a close by contemplating how one might go about turning the monstrousness approach into something more practicable. Recall that what this approach is supposed to measure is the confirmational (dis)jointedness between content parts of a hypothesis. Now, although we have limited access to the objective probabilities that determine the said confirmational (dis)jointedness we are not completely in the blind. One kind of information that is more readily accessible concerns the deductive relations between different content parts. For example, we can at least judge whether or not two content parts are *logically* independent. Moreover, we can judge whether or not any of their relevant (and non-redundancy containing) deductive consequences are logically independent. And we can even judge whether or not the conjunction of any pair of such consequences yields a relevant deductive element. Although these judgments fall short of empowering a full assessment of the monstrousness of a hypothesis, they at least give us some hints about the general trajectory such an assessment ought to take. Whether or not these hints are more informative than MDL-derived recommendations remains to be seen. My hunch is that, allowing for suitable modifications, these two approaches should in principle be capable of reaching similar levels of informativeness.

6. Conclusion
To summarise: In the first part of the essay I briefly considered some prominent philosophical accounts of ad hoc-ness and argued that these are deficient in some important respects. I then

---

[18] There are also intimate connections between the minimum description length approach and the idea of a content element, as it is conceived by Schurz and Weingartner (2010), since the latter is understood in terms of the *length* of a formula after it has been transformed to its negation-normal form.

made the case that appeal to intuitive judgments would only help adjudicate between rival conceptions of ad hoc-ness if those judgments were shared, something that is plainly not true. As an alternative, I recommended that we drop the intuitively-loaded term 'ad hoc-ness' and shift our focus onto a genuinely undesirable feature of hypotheses. I dubbed this feature that is often, but not universally, present in presumed examples of ad hoc-ness the 'monstrousness' of a hypothesis. I then proceeded to explicate this notion by means of the technical notions of probabilistic independence, content part, relevant consequence and relevant deductive element. I followed that up with a proposed measure of the monstrousness of hypotheses. I subsequently, and admittedly fleetingly, compared this monstrousness measure with the general spirit of MDL approaches to simplicity/ad hoc-ness. The outcome of that comparison was that the former profits from its a-posteriori attitude towards the problem it studies but only at the expense of practicability while the latter does the opposite. The essay concluded with a suggestion of how to understand monstrousness in a more practicable manner, one that hopefully leads to actionable advice concerning the conjecturing, extending and modifying of hypotheses.

Acknowledgements

References:

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.

Earman, J., & Glymour, C. (1978). Einstein and Hilbert: Two months in the history of General Relativity. *Archive for History of Exact Sciences*, 19, 291-308.

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science*, 45, 1-35.

Gemes, K. (1994). A new theory of content I: Basic content. *Journal of Philosophical Logic*, 23(6), 595-620.

Gemes, K. (1997). A new theory of content II: Model theory and some alternatives. *Journal of Philosophical Logic*, 26(4), 449-476.

Grünbaum, A. (1976). Ad hoc auxiliary hypotheses and falsificationism. *British Journal for the Philosophy of Science*, 27(4), 329-362.

Holton, G. (1969). Einstein, Michelson, and the "crucial" experiment. *Isis*, 60(2), 132-197.

Leplin, J. (1975). The concept of an ad hoc hypothesis. *Studies in History and Philosophy of Science*, 5(4), 309-345.

Popper, K. R. (1972). *Objective knowledge*: *An evolutionary approach.* Oxford: Clarendon Press.

Popper, K. R. (1974). Replies to my critics. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 961-1197). Open Court.

Redhead, M. (1978). Ad Hocness and the appraisal of theories. *British Journal for the Philosophy of Science*, 29(4), 355-361.

Rissanen, J. (1983). Universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11(2), 416-431.

Schurz, G. (1991). Relevant deduction: From solving paradoxes towards a general theory. *Erkenntnis*, 35, 391-437.

Schurz, G. (2014). Bayesian pseudo-confirmation, use-novelty, and genuine confirmation. *Studies in History and Philosophy of Science*, 45(1), 87-96.

Schurz, G., & Weingartner, P. (2010). Zwart and Franssen's impossibility theorem holds for possible-world-accounts but not for consequence-accounts to verisimilitude. *Synthese*, 172, 415-436.

Shavlik, J. W., & Dietterich, T. G. (Eds.) (1990). *Readings in machine learning*. San Mateo, CA: Morgan Kaufmann Publishers.

Votsis, I. (2014). Objectivity in confirmation: Post hoc monsters and novel predictions. *Studies in History and Philosophy of Science*, 45(1), 70-78.

Zahar, E. (1973). Why did Einstein's programme supersede Lorentz's? (Part I). *British Journal for the Philosophy of Science*, 24, 95-123.