

# Chapter 18

## Ad Hoc Hypotheses and the Monsters Within

Ioannis Votsis

**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 successfully 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 universally 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 somewhat sketchily compared with the minimum description length approach.

**Keywords** Ad hoc • Scientific methodology • Minimum description length • Philosophy of artificial intelligence • Computational science

---

I. Votsis (✉)  
New College of the Humanities, London  
e-mail: [ioannis.votsis@nchlondon.ac.uk](mailto:ioannis.votsis@nchlondon.ac.uk)

## 18.1 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 successfully 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 conceptions. After raising some problems specific to these conceptions, it is pointed out that there is a more general problem, 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 that is 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 somewhat sketchily 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 with respect to one way the latter is 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.

## 18.2 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".<sup>1</sup> 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. The result is that 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 two 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 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 hocness' which the scientific community employs" (p. 316) [original emphasis]. Why it is misguided will, I hope, become clear in the next section. The second contribution is Forster and Sober (1994). Here it is contended that the term 'ad hoc' signifies unreasonable revisions to scientific theories and, in particular, to auxiliary hypotheses. Though Forster and Sober do not attempt to pin down the notion, they allude to an important connection between non-ad hoc-ness and simplicity.<sup>2</sup> The suggestion, roughly, is that simpler hypotheses are less ad hoc. We return to this connection in Sect. 18.5 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

---

<sup>1</sup>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](http://www.oxforddictionaries.com/us/definition/american_english/ad-hoc?q=ad+hoc)

<sup>2</sup>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).

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  $H$ , which in this context is taken to be the explicans, 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 a non-ad hoc 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  $Z_I$  offers an ad hoc explanation of  $S_I$ .

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

$S_I$ : Sometimes the sky lights up with thunderbolts.

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

$A_I$ : Free falling objects near the earth's surface accelerate roughly at  $9.81 \text{ m/s}^2$ .

Following Popper's conception, conjunction  $Z_I \wedge A_I$  is not an ad hoc explanation of  $S_I$  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. But, surely, the conjunction is no less ad hoc than before. In other words, having 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.<sup>3</sup>

---

<sup>3</sup>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.

Consider next Zahar's approach. I here focus on one of three proposed conceptions put forth by him.<sup>4</sup> 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.<sup>5</sup> 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 (and where those phenomena were not novel for its predecessor) has excess testable content over and above 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.<sup>6</sup>

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. 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  $O_2$

---

<sup>4</sup>Here are the other two: "It is *ad hoc*<sub>2</sub> if none of its novel predictions have been actually 'verified' . . . [a] theory is said to be *ad hoc*<sub>3</sub> 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 hoc*<sub>1</sub> and his own notion ad hoc (c). For a more detailed discussion of Zahar's notions, particularly *ad hoc*<sub>2</sub>, the reader may consult Redhead (1978).

<sup>5</sup>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.

<sup>6</sup>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.

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.<sup>7</sup>

### 18.3 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 to do 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, 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 desirable. Secondly, and more importantly, it is highly doubtful that there is a uniform meaning of ad hoc-ness in the non-ordinary, i.e. 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 [is], 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 relevantly, 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

---

<sup>7</sup>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).

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.<sup>8</sup> 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 carryover 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. For reasons that will soon become apparent, I propose the use of the term ‘monstrousness’.<sup>9</sup>

## 18.4 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 thus 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.<sup>10</sup> Numerous examples of undesirable ad hoc hypotheses exhibit this disjointedness in good measure. This is certainly true of  $Z_I \wedge A_I$ . 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 (confirmational) disjointedness?<sup>11</sup> I propose the following articulation:

---

<sup>8</sup>One possible approach which is eschewed here, though it is legitimate, is to admit several distinct notions of ad hoc-ness.

<sup>9</sup>A version of this notion was first explored in Votsis (2014).

<sup>10</sup>For a discussion of the notion of content part you may consult Gemes (1994, 1997).

<sup>11</sup>Elsewhere, see Votsis (2015), I call this notion ‘confirmational disconnectedness’. Either name is fine.

*Disjointedness:* Any two content parts of a non-self-contradictory proposition  $\Gamma$  expressed as propositions  $A, B$  are disjointed if and only if for all pairs of internally non-superfluous and non-sub-atomic propositions  $\alpha, \beta$  where  $\alpha$  is a relevant deductive consequence of  $A$  and  $\beta$  is a relevant deductive consequence of  $B$ : (i) there is no true or partly true proposition  $\gamma$  such that  $\gamma$  is a relevant deductive consequence of  $\alpha$  and  $\gamma$  is also a relevant deductive consequence of  $\beta$ , (ii)  $P(\alpha/\beta) = P(\alpha)$  where  $0 < P(\alpha), P(\beta) < 1$  and (iii) there is no atomic proposition  $\delta$  that is a relevant deductive consequence of  $\alpha \wedge \beta$  and is not a relevant deductive consequence of either  $\alpha$  or  $\beta$  on their own.

Hence jointedness can be articulated thus:

*Jointedness:* Any two content parts of a non-self-contradictory proposition  $\Gamma$  expressed as propositions  $A, B$  are jointed if and only if for some pair of internally non-superfluous and non-sub-atomic propositions  $\alpha, \beta$  where  $\alpha$  is a relevant deductive consequence of  $A$  and  $\beta$  is a relevant deductive consequence of  $B$ : either (1) there is a true or partly true proposition  $\gamma$  such that  $\gamma$  is a relevant deductive consequence of  $\alpha$  and  $\gamma$  is also a relevant deductive consequence of  $\beta$ , or (2)  $P(\alpha/\beta) \neq P(\alpha)$  where  $0 < P(\alpha), P(\beta) < 1$  or (3) there is at least one atomic proposition  $\delta$  that is a relevant deductive consequence of  $\alpha \wedge \beta$  but is not a relevant deductive consequence of either  $\alpha$  or  $\beta$  on their own.

Let us take a closer look at the notion of disjointedness. What is a content part? Roughly put, a content part  $c$  of  $\Gamma$  is a non-trivial consequence of  $\Gamma$  from which one cannot derive  $\Gamma$  itself. In other words,  $c$  is strictly smaller in content than  $\Gamma$ . The non-superfluousness clause is there to remove superfluous content from the evaluation. One reason for doing so is that it reduces the evaluation's complexity. Note that the clause wards off superfluousness only within each of  $\alpha$  or  $\beta$ , but not across them, i.e. between  $\alpha$  and  $\beta$ . In short, the contents of  $\alpha$  and  $\beta$  are allowed to overlap. The reason we permit such overlap is that it allows the detection of one kind of jointedness.

Now consider clause (ii). 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  $\alpha, \beta$  are probabilistically independent we know that the probability of the one is not affected if we assume something about the truth (/falsity) of the other. 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.

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  $\alpha$  and a proposition  $\beta$ , 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  $\alpha_i, \beta_i$  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).<sup>12</sup>

Now consider clause (i). If there were such a proposition  $\gamma$ ,  $\alpha$  and  $\beta$  would share content and hence any confirmation of that content would confirm at least a content part of each  $\alpha, \beta$ . In short, such a proposition  $\gamma$  would force us to admit that  $\alpha$  and  $\beta$  and hence  $A$  and  $B$  are confirmationally linked. That’s why clause (i), *qua* a disjointedness clause, requires that there is no such proposition  $\gamma$ .

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  $\delta$  that neither entails on its own and whose truth would confirm both.<sup>13</sup> Clause (iii)

<sup>12</sup>There is also an analogous notion that applies to predicates – see Schurz (2014).

<sup>13</sup>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. 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

is there to ensure that there is no ‘indirect’ confirmational relation between  $A$  and  $B$  via such a consequence  $\delta$ . 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  $\delta$ , the concept of disjointedness is once again rendered unsatisfiable. That is, if we required only that there is no  $\delta$  that any  $\alpha, \beta$  jointly (but not individually) entail then there would almost always be such a  $\delta$ , more or less notwithstanding what content  $A, B$  possess. For example, one such  $\delta$  is  $\alpha \wedge \beta$  where  $\alpha$  is not logically equivalent to  $\beta$ . More generally, a proposition  $\zeta$  that is jointly (but not individually) entailed by two propositions  $\alpha, \beta$  cannot play the role of  $\delta$  if it is logically equivalent to  $\omega \wedge \varepsilon$  where both  $\omega$  is a relevant consequence of  $\alpha$  and  $\varepsilon$  is a relevant consequence of  $\beta$ . Such joint consequences are trivial for our purposes and therefore incapable of assisting us in our quest to find a confirmational relatedness between  $A, B$  that is not captured by conditions (i) and (ii). 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 notion of an atomic proposition. This not the same as an atomic sentence or proposition in the logical sense of the word. Rather, in the current sense (and roughly speaking) an atomic proposition contains the least amount of content that does empirically significant work. More formally, a proposition  $\varphi$  is atomic if, and only if,  $\varphi$  is non-superfluous and truthfully represents all and only the content of an atomic state of the world. Crucially, the content of an atomic proposition cannot be decomposed into distinct atomic or molecular content parts. How does it rule out guaranteed joint consequences like  $\alpha \wedge \beta$ ? If  $\delta$  is atomic and relevantly follows from  $\alpha \wedge \beta$ , then its potential decomposition into an equivalent conjunction could not be the conjunction  $\alpha \wedge \beta$ . For suppose  $\delta$  is equivalent to  $\alpha \wedge \beta$ . Recall that  $\alpha, \beta$  are (by stipulation) not sub-atomic. That means that either one or both are atomic or molecular in which case  $\delta$  is not atomic. Contradiction! In other words, requiring atomicity forbids such decompositions and hence consequences like  $\alpha \wedge \beta$  are ruled out.

What use could we possibly 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  $S_I$  confirms (a part of)  $Z_I$  precisely because  $Z_I$  was designed to entail  $S_I$  but, crucially, this confirmation does not spread to the non- $S_I$  part of  $Z_I$ , 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

---

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.

*genuine* confirmation when the support gained from some piece of evidence does not spread to content parts other than those corresponding to the evidence. My approach thus offers a unified treatment of confirmation relations.

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 monstrosity of a hypothesis. But before we do that we must consider one last complication.

Should judgments of monstrosity be affected by the way the same content is distributed between two different 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 monstrosity 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 monstrosity 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 monstrosity 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 monstrosity should ensure that  $w_1 = w_2$ . That is to say, monstrosity judgements should be invariant under different ways of distributing the same content between two different propositions.

One, perhaps the only, way to pull this off is to calculate monstrosity scores on the basis of *all* distinct ways of distributing the same content between two different 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.<sup>14</sup> Without further ado, here's the proposal: The monstrosity of a proposition is given by the ratio of the sum of disjointed pairs of parts taken from all distinct ways of distributing its content to the sum of the total number of pairs of parts (i.e. jointed and disjointed) taken from all distinct ways of distributing its content. Formally, the monstrosity  $m$  of a proposition  $\Delta$  is given by the following function:

$$m(\Delta) = \frac{\sum_{i=1}^n d_i^{\alpha,\beta}}{\sum_{i=1}^n t_i^{\alpha,\beta}}$$

---

<sup>14</sup>One reason for its inadequacy is that when a pair of propositions  $\alpha, \beta$  are deemed jointed the strength of their jointedness, e.g. the degrees of their probabilistic dependence, is neglected. Arguably such information should have a role in any suitably sensitive measure of monstrosity.

where  $d_i^{\alpha, \beta}$  denotes the number of disjointed pairs  $\alpha, \beta$  in a given content distribution  $i$ ,  $t_i^{\alpha, \beta}$  denotes the total number of jointed plus disjointed pairs  $\alpha, \beta$  in a given distribution  $i$ , and  $n$  denotes the total number of content distributions.<sup>15</sup>

The number of disjointed pairs in a given content distribution is determined by counting how many times a different pair of relevant deductive consequences  $\alpha, \beta$  turns out to satisfy clauses (i) - (iii) at the same time. Any pair that is not disjointed is counted as jointed. The higher (/lower) the value of  $m$  the more (/less) monstrous the content expressed by  $\Delta$ . Note that this value is the same no matter how we cut  $\Delta$  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 monstrosity 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.

## 18.5 Monstrosity 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 preferred hypothesis, according to this approach, is the one that provides the most economical description 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.

---

<sup>15</sup>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 atomic propositions.

... 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 monstrosity 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 monstrosity score *and* a low MDL simplicity score – in the latter case since its length is presumably no shorter than the combined length of the individual propositions it seeks to summarise.<sup>16</sup> 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, at least on a 'face value' reading of the MDL account, for the abovementioned reason but *does not* get a high monstrosity score. The reason for the latter is that the individual facts are systematically related via common relevant deductive consequences, e.g. the claim that there are at least two white swans. A corollary is that under the monstrosity approach, and contra (a 'face value' reading of) MDL, a conjunction of propositions expressing related facts gets the same score as a generalisation of them since they both possess the same content, provided the clause 'and nothing else satisfies the antecedent' is added into the mix, 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. Monstrosity 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 to be preferred 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 monstrosity 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 monstrosity approach into something more practicable. Recall

---

<sup>16</sup>There are also intimate connections between the minimum description length approach and Schurz's idea of a content element, since the latter is understood in terms of the *length* of a formula after it has been transformed to its negation-normal form.

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 atomic proposition. Although these judgments fall short of empowering a full assessment of the monstrosity 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.

## 18.6 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 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 ‘monstrosity’ of a hypothesis. I then proceeded to explicate this notion by means of the technical notions of probabilistic independence, content part, relevant consequence and atomic proposition. I followed that up with a proposed measure of the monstrosity of hypotheses. I subsequently, and admittedly fleetingly, compared the monstrosity 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 monstrosity in a more practicable manner, one that hopefully contains hints towards the potential automation of tasks that include the conjecturing, extending and modifying of hypotheses.

**Acknowledgements** My sincerest thanks to three anonymous referees as well as to my colleagues, Gerhard Schurz and Paul Thorn, for valuable feedback on the material presented in this essay. I acknowledge the German Research Foundation (Deutsche Forschungsgemeinschaft) for funding my research under project B4 of Collaborative Research Centre 991: The Structure of

Representations in Language, Cognition, and Science. Part of this essay has been written while working on the project ‘Aspects and Prospects of Realism in the Philosophy of Science and Mathematics’ (APRePoSMa) during a visiting fellowship at the University of Athens. The project and my visits are co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program ‘Education and Lifelong Learning’ of the National Strategic Reference Framework (NSRF)—Research Funding Program: THALIS—UOA.

## 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). La Salle: 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.
- Shavlik, J. W., & Dietterich, T. G. (Eds.). (1990). *Readings in machine learning*. San Mateo: 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.
- Votsis, I. (2015). Unification: Not just a thing of beauty. *Theoria: An International Journal for Theory, History and Foundations of Science*, 30(1), 97–114.
- Zahar, E. (1973). Why did Einstein’s programme supersede Lorentz’s? (Part I). *British Journal for the Philosophy of Science*, 24, 95–123.