DEUS EX MACHINA AND THE AESTHETICS OF PROOF

Alan J. Cain

Centro de Matemática
Universidade do Porto
Rua do Campo Alegre, 687
4169–007 Porto
Portugal

Email: ajcain@fc.up.pt

Unexpectedness and inevitability, two of the aesthetic qualities G.H. Hardy identified as being properties of beautiful proofs [Har40, §18], together seem paradoxical: how can something be seen as both unexpected and inevitable? Two possibilities are that the inevitability only becomes apparent in hindsight, or that the strategy of the proof is unexpected but, once chosen, proceeds inevitably. In this essay, I suggest a different solution: I argue that the literary concept of *deus ex machina* can be used to clarify the notion of inevitability in proof and reconcile it with unexpectedness.

Deus ex machina (literally, 'god from the machine', henceforth abbreviated to *deus*) refers to a type of plot device used to resolve a seemingly intractable situation. The term is derived from ancient Greek drama, where such a resolution might be effected by a god intervening, with the actor playing the god being lowered onto the stage by a crane (the 'machine'). It has come to mean any event in a story that resolves a situation but which does not fit with the internal framework of the plot [LE]. Aristotle is the earliest extant author to complain of the unsatisfactory nature of such a resolution [*Poetics*, 1454a33–b8], and this disdain has continued to the present day. Aristotle gives the example of how Medea, at the end of Euripides's play of the same name, is rescued from Jason's vengeance by being carried off to Athens in the chariot of the god Helios [*Medea*, l.1314]; until this point, the play is free from divine intervention.

NARRATIVE AND PROOF

Parallels between proof and narrative have been explored by Thomas, who argues that '[l]ogical consequence is the gripping analogue in mathematics of narrative consequence in fiction; all physical causes, personal intentions, and logical consequences in stories are mapped to implication in mathematics' [Tho02, p.45]. My intention is to focus here on the notion of *deus* in narrative, and argue for a parallel notion of *deus* in proof: inevitability, in the Hardian sense, can then be thought of as avoidance of *deus*.

In a narrative, the reader is presented with a place, a time, and some characters, and is told how the characters interact with each other and the world they inhabit. The reader gradually builds up a mental conception of the world and the characters' motives and personalities. A narrative, at its most basic, can be a bare enumeration of events, perhaps disconnected, that conveys only superficial information about the world of the narrative: telling what happened without giving the reader any inkling of why. A better, fuller, narrative allows the reader to build up a coherent mental picture and understand the events and the characters' actions.

In a proof, the reader is presented with some mathematical objects, and, through the reasoning in the proof, gradually builds up a mental conception of how these objects behave. Part of this mental conception is embodied in the theorem that is proven. A proof, at its most basic, can be a bare listing of statements, each following logically from earlier ones, that lead from the hypotheses to the conclusion. However, a better proof can leave the reader, not simply with knowledge of the theorem's truth, but with a deeper understanding of it and the objects it concerns.

Let me explain this further. In reading a narrative or a proof, a reader has some mental conception of the world of the narrative or the objects with which the proof is concerned. This conception includes *formal knowledge* and (for want of a better word) *intuition*. In narrative, formal knowledge consists of facts that are established within the world of the narrative. These may include facts about characters' past actions, skills, and relationships. In proof, it consists of definitions and proven properties of the objects. Intuition, in narrative, consists of less certain impressions, for instance regarding the motivation and psychology of the characters and expectations for how the plot will proceed. In a proof, intuition consists of an impression of how the objects concerned behave and interact.

In both cases, intuition is informed by formal knowledge. Indeed, the reader constructs an intuition of the characters of a narrative or the objects of a proof from those pieces of formal knowledge supplied by the narrative or proof. As a reader follows a proof or narrative, he acquires new formal knowledge. In a proof, readers will check each new step in the proof to see whether it follows from their existing stock of formal knowledge, and, if so, adds it to their formal knowledge. In a narrative, the checking is less important, or at least less active, although presumably a reader would notice if a narrative contained contradictory statements. In both narratives and proofs, the acquisition of new formal knowledge causes the reader to modify his intuition.

DEUS EX MACHINA

In a narrative, *deus* is unsatisfying for two reasons. The first is that it spoils any future attempt to build tension if the author has established that a difficulty can be resolved by a *deus*. The second, and for the purposes of this essay more important, reason is that it does not fit with the internal structure of the story. There is no reason *internal to the story* why the *deus* should intervene at that moment. There is only an external explanation: the author wants to extricate their hero. For this reason, readers cannot incorporate the *deus* and its consequences into the intuitive component of their mental conception, or at least can do so only with difficulty.

In the context of proofs, this second reason has a parallel. In a proof, *deus* takes the form of an unexplained construction or a calculation of elements or a definition of a function that simply 'happens to work'. Like *deus* in narratives, such a manoeuvre serves an external purpose, namely the teleological one of

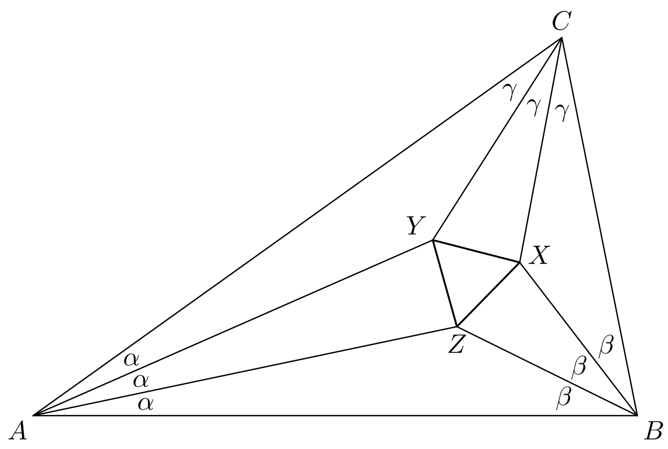

Figure 1: Morley's trisector theorem

proving the theorem at hand. Such features in a proof do not fit with the structure or setting of the proof. Readers cannot see why this construction or this calculation or that definition is being carried out; they cannot perceive a reason for it that is internal to the proof. In short, they have more difficulty in modifying their intuitive conceptions to include the *deus*. They can follow the proof to its conclusion, checking each step against their formal knowledge of the objects concerned, but the *deus* is a cataract that their intuition cannot easily navigate.

I argue that taking 'inevitability' in proof to mean 'avoidance of *deus*' allows one to understand how 'inevitability' and 'unexpectedness' can both occur in a beautiful proof: for like a pleasing narrative, a beautiful proof can contain unexpectedness provided it fits its structure and setting.

A case study: three proofs of Morley's theorem

To illustrate the notion of *deus* in proof, I shall compare three different proofs of Morley's trisector theorem, which says that the adjacent trisectors of the angles of any triangle meet at the vertices of an equilateral triangle; see Figure 1. [There are many different proofs of this theorem; see the bibliography in [OB78].] I shall not gives the proofs in full, I shall merely highlight the salient points.

1. Conway's proof [Con] starts with three angles α, β, γ with $\alpha + \beta + \gamma = \pi/3$. It specifies the angles and side lengths of seven triangles, one being equilateral, and then shows that they can be fitted together to form the triangles ABZ, ACY, BCX, AYZ, BXZ, CXY, and XYZ in Figure 1. For example, one triangle (which will become AYZ when the pieces are assembled) is specified to have angles α, $\gamma + \pi/3$, $\beta + \pi/3$, with the side between the latter two angles being equal to the side of the equilateral triangle. Thus the proof constructs the triangle ABC.

2. The proof given by Dijkstra [Dij82, p.182–3] again starts from three angles α, β, γ. It starts from the equilateral triangle XYZ and constructs the triangles AYZ, BXZ, and CXY, with the angles of each being specified. For example, AYZ is specified to have angles α, $\gamma + \pi/3$, $\beta + \pi/3$. It

then uses the sine rule to show that $\alpha = \angle CAY = \angle YAZ = \angle BAZ$, and similarly for the angles at B and C.

3. The proof by Bankoff [Ban62] starts from the triangle ABC and its trisectors, and, by making use of trigonometric identities and the sine rule, first of all calculates the lengths of AY and AZ in terms of the radius of the circumcircle and the angles of ABC, then calculates the angles $\angle AYZ$ and $\angle AZY$. Symmetrical arguments give the angles $\angle BXZ$, $\angle BZX$, $\angle CXY$ and $\angle CYX$, from which it follows that each angle of XYZ must be $\pi/3$.

Conway's proof (1) is the shortest of the three. It is simple and has the merit of avoiding use of trigonometric identities or the sine rule, but the specification of the seven triangles is *deus*. The values for what turn out to be the angles of the seven smaller triangles of Figure 1 simply happen to work. Certainly, one sees some of the relationship between the angles, such as the fact that $\angle AYZ$ is dependent on γ. But one does not see why it is dependent only on γ and not on α or β or the side length of the triangle.

Dijkstra's proof (2) also involves a *deus*, albeit a milder one: only the specification for three triangles is produced out of a hat. The use of the sine law then gives the reader some intuitive feeling of how the result follows from the relationships between the sides and angles of the triangle.

Bankoff's (3) is the longest proof, but its approach is unsurprising. It uses the kind of trigonometric arguments one expects in this situation, including several applications of the sine rule. Although this argument is rather more involved than either of the other two, requiring the use of trigonometric identities and several applications of the sine rule, it contains no *deus*. In particular, the reader can follow the reasoning intuitively.

Both Conway's and Dijkstra's proofs work in reverse: they start from the equilateral triangle XYZ and show that for any angles α, β, γ, a triangle with angles 3α, 3β, 3γ can be constructed whose adjacent angle trisectors meet at X, Y, and Z. Bankoff's, in contrast, starts from the triangle ABC and deduces that XYZ is equilateral. It gives an idea of the relationships holding between the angle trisectors and how these force XYZ to be equilateral. For example, one can see why $\angle AYZ$ is dependent only on γ: because AY and AZ are also dependent on α, β, and the radius of the circumcircle, but these dependencies cancel each other out.

Comparison of these examples also shows the independence of economy (the third of Hardy's aesthetic qualities of beautiful proofs) from the absence of *deus*. For Conway's proof, with the strongest *deus*, is the most economical. The only tools it uses are the most elementary geometric facts; the most 'advanced' being the 'angle–side–angle' similarity argument. Dijkstra's has a milder *deus* but uses a more advanced tool, viz., the sine rule. Bankoff's avoids *deus* but requires a still bigger toolkit: the sine rule and various trigonometic identities.

DIFFERENCES BETWEEN NARRATIVE AND PROOF

One difference between *deus* in narrative and *deus* in proof should be emphasized. Chronology, in the sense of the order in which a reader is informed of events, plays an important rôle in narrative. Whether a reader views a particular event as *deus* is closely linked with chronology. If the readers have been supplied with an explanation for an event (in the sense of knowledge of causes for this event or at least the potentiality of this event), they will not view the event as *deus*. This holds true even if the explanation is not recognized as such until the reader encounters the event. (The mysterious figure who has been following the hero for days is revealed to be an ally and comes to his aid.) However, an event followed by a *post hoc* explanation is unlikely to be appreciated. (The hero is rescued by an ally of whom the reader had not been hitherto informed, but who is now said to have been following him for days.) The readers may be able to incorporate the *post hoc* explanation into their intuition, but the author is unlikely to be able to salvage their ability to create tension in the remainder of the narrative.

In proof, chronology plays a lesser rôle, at least insofar as the evaluation of *deus* is concerned. If, after a construction or calculation or definition, the proof retrospectively shows why this procedure was necessary, readers will be able to incorporate it into their intuition. The most basic example of this is probably Euclid's proof of the infinity of prime numbers [*Elements* IX.20]. Given a collection of primes $p_1, \ldots, p_n$, one forms the number $N = p_1 \cdots p_n + 1$. Only retrospectively does the reason for this definition for N become clear: because division of N by any p_i leaves remainder 1, so that some prime not among the p_i must divide N. This retrospective explanation for the choice of N ensures that this move is not seen as *deus*; indeed Hardy selects Euclid's proof as one of his two examples of beautiful proofs [Har40, §12]. In contrast, in Conway's proof of Morley's theorem, for instance, the specification of angles is not justified retrospectively.

INEVITABILITY AND UNEXPECTEDNESS

In light of these discussions, the Hardian aesthetic concept of inevitability in proof can be seen as avoidance of *deus*. Hence, unexpectedness can be reconciled with inevitability, for avoiding *deus* does not entail avoiding unexpectedness.

Just as a subject, location, character, or starting-point for a narrative can each be unexpected without being a *deus*, so an approach to a proof can be surprising and yet, once the approach is selected, the proof can proceed inevitably to its conclusion without invoking *deus* at any point. For example Zagier's one-sentence proof of Fermat's theorem that every prime $p \equiv 1 \pmod 4$ is a sum of two squares [Zag90] has an unexpected starting point, but thenceforth the proof proceeds inexorably (and rapidly) to its conclusion.

An event in a story can be unexpected yet not a *deus* if, once the reader has encountered it, it is seen to fit properly with the structure of the tale. The same holds in proof: Netz [Net05, p.256–7] noted an example in Archimedes's *Sphere and Cylinder*. Archimedes announces his intention of proving decidedly three-dimensional results: that the surface of a sphere is four times its great circle, and that the volume of a sphere is two-thirds of a cylinder that exactly encloses it. Yet he starts by proving two-dimensional results: a series of theorems on circles and polygons. This is followed by results on pyramids and cones; three-dimensional results, true, but apparently irrelevant. Then, suddenly, he imagines rotating polygons about an axis: all the two-dimensional results suddenly acquire three-dimensional analogues and their applicability to the situation at hand suddenly makes sense to the reader; the desired results follow in short order. This is unexpected, but not a *deus*, for it fits with the internal structure of the proof and illuminates the preliminary results.

OTHER EXAMPLES

Einstein discussed what he perceived as an ugly and an elegant proof of one direction of Menelaus's theorem. [Given a triangle ABC and a line dividing the lines AB, BC, and CA into $A'B$ and $A'C$, $B'C$ and $B'A$, $C'A$ and $C'B$, respectively, $A'C \cdot B'A \cdot C'B = A'B \cdot B'C \cdot C'B$.]

> 'Although the first proof is somewhat simpler, it is not satisfying. For it uses an auxiliary line that has nothing to do with the content of the proposition to be proved, and the proof favors, for no reason, the vertex A, although the proposition is symmetrical in relation to A, B, and C. The second proof, however, is symmetrical, and can be read off directly from the figure.' [LL90, p.38]

In the terms used in this paper, the auxiliary line and the favouring of the vertex A form a *deus*, for nothing either before or afterwards compels these constructions.

Some mathematicans have a certain distaste for the Haken–Appel proof of the four-colour theorem, which depends upon a computer-assisted argument (see, for example, [Kin06, p.92–3]). The computer-assisted part of the proof is a *deus*, albeit of a slightly different kind to the examples discussed above. For in the earlier examples, the *deus* still allows a reader to formally check the proof, whereas the appeal to a lemma proved by a computer does not allow this checking. Even leaving aside such issues of validity or surveyability (see, for instance, [Tym79]), the computer-assisted part of the proof seems to be a *deus*, for it represents a step that readers cannot easily incorporate into their intuition. Certainly one can see the strategy of that part of the proof: that all of the unavoidable configurations are reducible. But the proof for this is too long for readers to follow so that they can gradually modify their intuition; the reader is essentially forced to jump ahead in the proof and simply asked to accept the correctness of intermediate steps.

Imagine a hypothetical narrative parallel: a variation of book IX of the *Odyssey* wherein Odysseus tells Alcinous of his wanderings since the fall of Troy, reaches the point when he and his men are trapped by the Cyclops Polyphemus, but then jumps ahead to leaving the island, and blandly assures his listeners that his cunning allowed their escape. We would have reason to believe him: he successfully uses stratagems and ruses at many points throughout the *Odyssey*. Yet, even if we were to accept the truth (within the world of the tale) of what he says, such a turn of events would be a *deus*, for we could not modify our intuition to take us from the situation of their being in Polyphemus's power to their leaving the island. Our intuition would be that they are in an inescapable situation. To suddenly jump ahead to after their escape would be just as aesthetically unsatisfying as, for instance, Polyphemus spontaneously deciding to release them. Unable to modify our intuition, we would be left with unresolved questions: How could they overcome Polyphemus's great strength? How could they evade the other Cyclopes? The real story, by contrast, tells how Odysseus and his men prepare their plan; how Odysseus creates an opportunity to use it by giving Polyphemus wine; and how he sets up their escape past the other Cyclopes. Each step here allows us to modify gradually our intuition of Odysseus and of Polyphemus.

connections with teaching and exposition

Aside from aesthetics, various authors have argued that in expositional work, one should avoid manoeuvres that are akin to *deus* as I use the term: Chow says that 'every step should be motivated and clear' [Cho09, p.1] and follows Newman in saying that proofs should be 'natural' in 'not having any ad hoc constructions or *brilliancies*' [New98, p.59, italics in original]. Tucker explicitly recommends that, when teaching the calculus, one should not use 'deus-ex-machina auxiliary functions' [Tuc97, pp.239–240].

This expositional advice, if followed strictly, would seem to rule out the use of surprise, whereas I have drawn a distinction between *deus* (in my sense) and unexpectedness in proofs. Certainly, it seems pedagogically safer to avoid both and, following Chow's advice, motivate every step. Additionally, a proof is probably easier to memorize if it avoids *deus*, for each *deus*, not being compelled by the overall structure of the proof, would have to be explicitly remembered. However, complete avoidance of surprise might reduce the appeal of the exposited mathematics.

aesthetics of proof

Proofs, like narratives, can be aesthetically unsatisfying in ways other than using *deus*. Just as a narrative text can use inelegant language, clumsy exposition, or bad pacing, a proof can use poor notation, unclear explanation, or unsatisfactory division into lemmata. Each of these factors would decrease the satisfaction of a narrative or a proof that nevertheless avoided *deus*.

Rota [Rot97, p.181] suggests that 'mathematical beauty' is a term mathematicians use to avoid describing a piece of mathematics as enlightening. One does not need to accept fully his assertion to see that the identification of beauty with enlightenment is compatible with the arguments above. As readers follow a proof, they modify their intuition. They will find the proof enlightening if, by the end of the proof, their intuition includes what the theorem describes. If the readers' intuition does not include it, the proof is unenlightening. Since *deus* presents the readers with a difficulty in modifying their intuition, a proof that involves *deus* is less likely to be perceived as enlightening.

acknowledgements. The author thanks Yumi Murayama for reading and commenting on an earlier draft of this essay.

references

[Ban62] L. Bankoff. A simple proof of the Morley theorem. *Mathematics Magazine*, 35(4):223–224, 9 1962.

[Cho09] T. Y. Chow. A beginner's guide to forcing. In T. Y. Chow and D. C. Isaksen, editors, *Communicating Mathematics*, volume 479 of *Contemporary Mathematics*, pages 25–40. American Mathematical Society, 2009.

[Con] J. H. Conway. Posted to the geometry.puzzles newsgroup, 24th November 1997.

[Dij82] E. W. Dijkstra. A collection of beautiful proofs. In *Selected Writings on Computing: A Personal Perspective*, Texts and Monographs in Computer Science, pages 174–183. Springer-Verlag, New York, 1982.

[Har40] G. H. Hardy. *A Mathematician's Apology*. Cambridge University Press, Cambridge, 1940.

[Kin06] J. P. King. *The Art of Mathematics*. Dover, Mineola, 2006.

[LE] Deus ex machina. The Literary Encyclopedia. 28 June 2004. [http://www.litencyc.com/php/stopics.php?rec=true&UID=1436, accessed 17 July 2009.].

[LL90] A. S. Luchins and E. H. Luchins. The Einstein-Wertheimer correspondence on geometric proofs and mathematical puzzles. *Mathematical Intelligencer*, 12(2):35–43, 1990.

[MJP05] P. Mancosu, K. F. Jørgensen, and S. A. Pedersen, editors. *Visualization, Explanation and Reasoning Styles in Mathematics*, volume 327 of *Synthese Library*, Dordrecht, 2005. Springer.

[Net05] R. Netz. The aesthetics of mathematics. In Mancosu et al. [MJP05], pages 251–293.

[New98] D. J. Newman. *Analytic Number Theory*, volume 177 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.

[OB78] C. O. Oakley and J. C. Baker. The Morley trisector theorem. *American Mathematical Monthly*, 85(9):737–745, 1978.

[Rot97] G. C. Rota. The phenomenology of mathematical beauty. *Synthese*, 111(2):171–182, 5 1997.

[Tho02] R. S. D. Thomas. Mathematics and narrative. *Mathematical Intelligencer*, 24(3):43–46, 2002.

[Tuc97] T. W. Tucker. Rethinking rigor in calculus: the role of the mean value theorem. *American Mathematical Monthly*, 104(3):231–240, 1997.

[Tym79] T. Tymoczko. The four-color problem and its philosophical significance. *Journal of Philosophy*, 76(2):57–83, 2 1979.

[Zag90] D. Zagier. A one-sentence proof that every prime $p \equiv 1 \pmod 4$ is a sum of two squares. *American Mathematical Monthly*, 97(2):144, 1990.