Generalisations, Examples and Counter-examples in Analysis and Optimisation
In honour of Michel Théra at 70

Jonathan M. Borwein FRSC, FAA, FAAAS, FAMS

CARMA, University of Newcastle NSW 2308

July 20, 2016

Abstract

In this essay, I talk about the role of examples and counter-examples and of generalisation in mathematical research. I use geometric fixed point theory and nonsmooth optimisation to illustrate my opinions and conclude with a few recommendations.

1 Introduction

Richard Brown discussing constructivism and intuitionism [16, p. 239] writes

> Philosophical theses may still be churned out about it,
>
> ...
>
> but the question of nonconstructive existence proofs or the heinous sins committed with the axiom of choice arouses little interest in the average mathematician. Like Ol' Man River, mathematics just keeps rolling along and produces at an accelerating rate "200,000 mathematical theorems of the traditional handcrafted variety ... annually." Although sometimes proofs can be mistaken—sometimes spectacularly—and it is a matter of contention as to what exactly a "proof" is—there is absolutely no doubt that the bulk of this output is correct (though probably uninteresting) mathematics." (Richard C. Brown [16])

A forty year professional life as a mathematics teacher, researcher and editor leads me to write this short essay. After decades of editing and even more of authoring and participation in conferences or workshops, I can no longer resist making some observations. View these as *advice to a young mathematician (author, referee, or editor)* but since life expectancy is still increasing at roughly three years a decade we are nearly all young longer.

2010 *Mathematics subject classification.* Primary 47H09, 90C25; Secondary 49J52

Key words and phrases. generalisation and unification, examples and counter-examples, geometric fixed point theory, non-smooth analysis.

Two interwoven topics exercise me. *What is a good generalisation and what role do examples play in illustrating ones work*?

In Brown's accurate – if somewhat gloomy – assessment of current mathematics generalisations of varying qualities are the daily work of most research mathematicians. No one sits down to write a bad or dull paper. Nonetheless many papers are at least the latter if not the former. Since most professional mathematicians have to publish, my aim is to mitigate the situation not to expunge the literature.

2 Generalisations: good, bad and indifferent

> "This skyhook-skyscraper construction of science from the roof down to the yet unconstructed foundations was possible because the behaviour of the system at each level depended only on a very approximate, simplified, abstracted characterization at the level beneath.[1] This is lucky, else the safety of bridges and airplanes might depend on the correctness of the "Eightfold Way" of looking at elementary particles.
>
> [1] ... More than fifty years ago Bertrand Russell made the same point about the architecture of mathematics. See the "Preface" to Principia Mathematica "... *the chief reason in favour of any theory on the principles of mathematics must always be inductive*, i.e., it must lie in the fact that the theory in question allows us to deduce ordinary mathematics. In mathematics, the greatest degree of self-evidence is usually not to be found quite at the beginning, but at some latter point; hence the early deductions, until they reach this point, give reason rather for believing the premises because true consequences follow from them, than for believing the consequences because they follow from the premises." Contemporary preferences for deductive formalisms frequently blind us to this important fact, which is no less true today than it was in 1910." (Herbert A. Simon (1916–2001) [32, p.16])

It is very pleasing that Noble economist Simon and the great deductive logician Bertrand Russell so clearly acknowledge that we first recognise progress in mathematics by the new theory's ability to recapture the known. In our papers we must remain aware of this situation.

Let me make some terms explicit. A *generalisation* is an extension of a known result to cover more cases (a unification), and/or weaken hypotheses and/or strengthen conclusions. The role of *examples and counter-examples* is to validate and substantiate one or more of these three roles.

A result of the form "Assume a,b, or c. Then A,B, or C follows" is not a true unification if – as often the case – the proof of each case is then as much work or more than that of the original. It is also not needed if the average researcher would find the extension easily and could observe that the original proof still works with minor and fairly obvious tweaks. The situation is not improved by the assertion that it captures the results by Professors α, β, or γ unless those results are worth recapturing.

I will next illustrate generalisations with various results in two areas I know well: (a) geometric fixed point theory and (b) nonsmooth optimisation. In the following section I will then revisit corresponding examples for these results.

2.1 Geometric fixed point theory

A lovely and deservedly well known theorem was introduced by Banach in his 1920 dissertation and published in [4] two years later. Its applications are ubiquitous and important. Almost all undergraduate mathematicians will have seen contraction mappings or cognate tools applied to establish implicit function theorems, Newton-type theorems, or the existence of local solutions to ordinary differential equations with given initial conditions.

A *contraction mapping* is a Lipschitz mapping with Lipschitz constant strictly less than one.

Theorem 2.1 (Banach contraction [4])**.** *Let (X,d) be a complete metric space. Suppose $F : X \to X$ is a contraction mapping:*

$$d(F(x), F(y)) \leqslant \alpha d(x, y)$$

for all $x, y \in X$. Then F has a unique fixed point ($x = F(x)$).

In the late 1960s, Nadler considered an extension to multivalued mappings with closed non-empty images.

Theorem 2.2 (Multivalued contraction [24])**.** *Let (X,d) be a complete metric space. If $F : X \to CB(X)$ (the closed bounded non-empty sets in X) is a multivalued contraction mapping (in the Hausdorff metric), then F has a fixed point ($x \in F(x)$).*

Among other things this extension may be seen as a precursor for application to iterated function systems yielding deterministic fractals as fixed point sets – as first discovered by Hutchinson [5, 15, §3,5]. The proof of Theorem 2.2 follows that of Theorem 2.1. A fine recent survey is provided in the monograph [15].

A *non-expansive mapping* is a Lipschitz mapping with Lipschitz constant less than or equal to one. A special case of a famous theorem independently proven by the three named authors is:

Theorem 2.3 (Browder-Göhde-Kirk [17])**.** *Let $(B_X, \|\cdot\|)$ be the closed unit ball in a uniformly convex Banach space. If $F : B_X \to B_X$ is a non-expansive mapping, then F has a fixed point ($x = F(x)$).*

In Hilbert space this result can be proven by considering the contraction $F_\lambda(x) := \lambda F(x) + (1-\lambda)x_0$ for some fixed $x_0 \in B_X$) and for $0 < \lambda < 1$ and then letting λ tend to one. In infinite dimensions this approach also relies on $I - F$ being maximal monotone and so *demi-closed* [13]. This approach does not work as easily in a more general setting. It is not known if this result is true in all reflexive spaces or even in all equivalent renormings of Hilbert space.

Nearing the other end of the spectrum is a result we quote from [21, 27, 28].

Theorem 2.4 (Kannan contraction)**.** *Let (X,d) be a complete metric space. Suppose for some $0 \leqslant \alpha < 1/2$ the function $F : X \to X$ satisfies*

$$d(F(x), F(y)) \leqslant \alpha[d(x, F(x)) + d(y, F(y))]$$

for all x, y in X, then F has a unique fixed point ($x = F(x)$).

Theorem 2.4 is independent of the contraction principle. We shall consider Kannan's examples in Section 3.1. This result and its peers are not without their uses – though they are few and far between [17]. My old colleague and mentor Michael Edelstein wrote the review MR043343 of [27]:

From the author's summary: "A number of authors have defined contractive type mappings on a complete metric space X which are generalizations of the well-known Banach contraction, and which have the property that each such mapping has a unique fixed point. In this paper we compare this multitude of definitions."

Some 250 definitions—together with their variants—make up this vast multitude.

Reviewed by M. Edelstein

A quarter of a century later the multitude is vaster and the utility little greater.

2.2 Nonsmooth critical points

A work-horse of classical [2][1] and modern numerical optimization [25] is the method of steepest descent. Going back to Cauchy, it is based on the following theorem. We recall that if E is a Euclidean space and $f\colon E \to \mathbf{R}$, the *gradient* of f at x, $\nabla f(x)$ is the vector of first partial derivatives $\nabla f(x) = [f_1(x), f_2(x), \ldots, f_n(x)]$. We also recall that the directional derivative of f at x in direction h is defined by

$$D_h f(x) = \lim_{0<t\to 0} \frac{f(x+th)-f(x)}{t} \tag{1}$$

when that limit exists.

Theorem 2.5 (Cauchy)**.** *Let E be a Euclidean space. Suppose that $f\colon E \to \mathbf{R}$ has continuous first partials around x. Then the directional derivative of f at x in direction h satisfies*

$$D_h f(x) = \langle \nabla f(x), h\rangle. \tag{2}$$

In particular, f is differentiable at x.

Roughly one hundred and fifty years later it was recognised that for a continuous, hence locally Lipschitz, convex function this directional derivative always exists and is sublinear in h. Hence if we define the *convex subgradient* or *subdifferential* by

$$\partial f(x) = \{y \in E\colon f(x+h) \geqslant f(x) + \langle y, h\rangle, \forall h \in E\} \tag{3}$$

we have a wonderful "max formula" generalising and improving Theorem 2.5 (in the convex case).

Theorem 2.6 (Moreau-Rockafellar, [11])**.** *Let $f\colon E \to \mathbf{R}$ be a continuous convex function*

$$D_h f(x) = \sup_{y\in\partial f(x)} \langle y, h\rangle \tag{4}$$

for all $h \in E$. In particular $\partial f(x)$ is nonempty.

In his thesis Francis Clarke extended this result to all locally Lipschitz functions. Clarke replaced $D_h f(x)$ by

$$D_h^c f(x) = \limsup_{0<t\to 0, y\to x} \frac{f(y+th)-f(y)}{t}. \tag{5}$$

[1]This is Armijo's only publication listed in Math. Reviews!

Somewhat miraculously the mapping p sending $h \to D^c_h f(x)$ is always continuous and sublinear in h, and so if we define $\partial^c f(x) = \partial p(0) = \{y \in X^*\colon \langle y, h\rangle \leqslant D^c_h f(x), \forall h \in X\}$, Theorem 2.6 leads directly to:

Theorem 2.7 (Clarke [18, 11, 26])**.** *Let $f\colon E \to \mathbf{R}$ be a locally Lipschitz function*

$$D^c_h f(x) = \sup_{y \in \partial^c f(x)} \langle y, h\rangle \tag{6}$$

for all $h \in E$. In particular $\partial^c f(x)$ is nonempty. Moreover, $\partial^c f(x)$ is a singleton if and only if f is strictly differentiable at x.

In truth Clarke, appealing to Rademacher's theorem, originally defined $\partial^c f(x)$ as the closed convex hull of limits of nearby points of differentiability. This makes (6) seem even more remarkable. There is, however, a dark side to the situation [12, Cor. 9]. Recall that a set in a Banach space is *generic* if it contains intersection of countably many dense open sets. The complement is thus very *small* topologically [13].

Theorem 2.8 (Generic triviality [12])**.** *Let A be an open subset of a Banach space X. Then the set of non-expansive functions on A with $\partial^c f(x) \equiv B_{X*}$ for all x in A is generic in the uniform norm on A.*

In other words, in the sense of Baire category the Clarke subdifferential (likewise the limiting subdifferential in the separable case) of almost all functions contains no information at any point of A. Of course none of these functions are convex since then the function is generically strictly Fréchet differentiable [13]. This is the cost of abstraction — if a construction always works for a very broad class it usually works only passingly well.

So, for most Lipschitz functions the Clarke calculus is vacuous. That is why serious researchers work with well structured subclasses such *semi-algebraic, partially smooth* or *essentially smooth* functions. The situation becomes even worse when the generalisation is poorly thought through or unnatural.

Let us turn to a less edifying generalisation of convexity: the notion of *invexity.* The original definition entails a relaxation of the gradient inequality for a differentiable convex function replacing $y - x$ by an arbitrary vector $\eta(x, y)$. The notion leads to the following very easy theorem originally due to Hansen.

Theorem 2.9 (Invexity [7])**.** *Let Let $f\colon E \to \mathbf{R}$ be a (Gateaux) differentiable function. Then f is invex if and only if every critical point is a global minimum.*

While pretty, this result should have ended the matter. The notion is so broad that only relatively trivial results can be deduced and there is certainly no useful calculus.

Instead it has spawned a host of trivial results, generalisations and extensions. It is immediate that *the family of functions invex with respect to the same η is a convex cone.* See also [23]. I challenge the reader to find a single result that does not follow from Theorem 2.9 and the preceding information. I am not the first person to rail against invexity [33]. Nor is it by a long shot the only such topic where notational generalisations are studied in vacuo. The sad thing is that in a closed sub-community, potentially good researchers can work hard to produce relative trivialities.

3 Examples: good, bad and indifferent

> The mathematician's patterns, like the painter's or the poet's must be beautiful; the ideas like the colours or the words, must fit together in a harmonious way. Beauty is the first test: there is no permanent place in the world for ugly mathematics. – G.H. Hardy (1887-1977) [19]

Beauty and good taste in mathematics are frequently discussed but rarely captured precisely. Terms like 'economy', 'elegance', 'naturalness', and 'unexpectedness' abound but for the most part a research mathematician will say "I know it when I see it" as with US Supreme Court Justice Potter Stewart's famous 1964 observation on pornography [9, p. 1]. But we can do a bit better than that.

When I was young, I disparaged the role of mathematical taste as too subjective and too idiosyncratic to be used to determine what should be published. Now I believe that even if we can not agree, the notions should be more openly discussed. Almost all mathematicians agree with Hardy until asked to put flesh on the bones of his endorsement of beauty. Beauty may be the first test, but it is in the eye of the beholder.

Hardy, in the twelfth of his twelve lectures given as a eulogy for the singular Indian genius Srinivasa Ramanujan (1887–1921), described a result of Ramanujan – now viewed as one his finest – somewhat dismissively as "a remarkable formula with many parameters."[2] mist Realising the value of a generalisation may thus take years or even decades. This is not an invitation to publish any and all generalisations. If any competent researcher could with little effort prove the result if needed, it should be left as a remark in a more substantial piece of work. I quite often do things like this in a probably vain attempt to limit such future publications. Of course you need to have checked carefully that it is an easy extension.

Correspondingly a deep result may not have any examples but its logic and relation to its precursors must be honestly discussed. For example, an early version of the celebrated Bishop–Phelps theorem first appeared under the title "A proof that every Banach space is subreflexive." (During the 1950's Phelps had studied subreflexive normed spaces.) As another example there is now a large literature lifting results in non-smooth optimisation from spaces with equivalent Fréchet smooth renorms to Asplund spaces [13]. These results typically proceed by *separable reduction* because each separable subspace of an Asplund space does possess a Fréchet renorm. The details are often technically impressive and so many of these results are well worth publish as *pure* functional analysis. There are examples of Asplund spaces without even a Gateaux renorm, but no optimiser will ever need to work in such a setting. So no such paper can be justified on the basis of its *utility*.

Let me distinguish two forms of example – contrived and natural. A *contrived example* is a single function or object designed to show that the promised generalisation is indeed logically properly more inclusive. A *natural example* introduces some qualitative property that plays not only a logical role but also a motivational role shedding light on the situation. Again à la Judge Potter Stewart we usually know the difference when we see it. Especially if we ask the right question.

A contrived example is better than none but not much. It does not provide adequate justification or motivation. I remember struggling with the elliptic PDE literature of half a century ago trying to understand why the specified equation needed to be studied. This may largely be because of my ignorance but many eminent authors offered little enlightenment. Another common and frustrating

[2] See Ole Warnaar's 2013 contribution in "Srinivasa Ramanujan Going Strong at 125, Part II," available at http://www.ams.org/notices/201301/rnoti-p10.pdf.

situation occurs when the authors of a generalisation blithely asserts that a very serious looking application is covered by their work – but give no details or so few as to leave the reader helpless.

3.1 Fixed Point Examples

Having already commented on the ubiquity of Banach's contraction method let me turn to the three generalisations of Section 2.1. One attraction of all fixed point methods is that a fixed point of $(1-\lambda)I + \lambda F$ produces a zero of F.

Example 3.1 (Multivalued contractions)**.** Nadler [24] gives many structure theorems showing that union and composition of multivalued contractions will remain in the class. While these are interesting, the example of the multivalued contraction on the unit circle $\mathbf{S}$ given by

$$F\colon e^{i\theta} \in \mathbf{S} \to \{-\mathbf{e}^{\mathbf{i}\theta/\mathbf{2}}, \mathbf{e}^{\mathbf{i}\theta/\mathbf{2}}\}$$

clinches the deal. It has a fixed point $1 \in F(1)$. In this case there is no continuous, and hence no contractive, selection of F. But F has a multivalued contraction constant of $1/\sqrt{2}$. Since Nadler gives no details — he makes a one sentence comment on the whole example — let me note that this reduces to showing that

$$\frac{H^2\left(\{1,-1\},\{e^{i\theta/2},-e^{i\theta/2}\}\right)}{d^2(1,e^{i\theta})} = \frac{\min(1 \pm \cos(\theta/2))}{\sin^2(\theta/2)}$$

which has a maximum of $1/2$; so F is a set-valued α-contraction for any $\alpha \geqslant 1/\sqrt{2}$.

Note that by Michael's selection theorem [11, 13], multivalued contractions with compact convex non-empty images always have continuous selections. ⋄

Example 3.2 (Nonexpansive mappings)**.** Translations, rotations, metric projections and reflections are all nonexpansive but are not typically contractions. The class is closed under projections. This surely provides a cornucopia of natural examples. Moreover, important applications abound [6] both theoretical and practical. ⋄

Example 3.3 (Kannan mappings)**.** Kannan in [21] gives three examples delimiting his result from Banach's. To be fair Kannan provides a more general result and he advertises it in his third example as applying in incomplete spaces.

Since the two of his three examples given below in sum show that the two results are logically incomparable Theorem 2.4 is more truly a *variation* than a generalisation. His two examples are

(a) Let $X = [0,1]$ in the usual metric and let $F(x) = x/4$ for $0 \leqslant x < 1/2$ and $F(x) = x/5$ for $1/2 \leqslant x \leqslant 1$. Then F is discontinuous at $1/2$ and so not a Banach contraction. However, Theorem 2.4 applies with $\alpha = 4/9$. This I would categorise as a contrived example. Indeed F need not be continuous at except at the fixed point. Consider, $F(x) = x/3$ when $x \in \mathbf{Q}$ and $F(x) = 0$ when $x \notin \mathbf{Q}$.

(b) Again let $X = [0,1]$ in the usual metric and let $F(x) = x/3$ for $0 \leqslant x \leqslant 1$.[3] Then F is a Banach contraction. However, Theorem 2.4 never applies as is seen by considering $x = 1/3, y = 0$. If not contrived this is a depressing example.

[3] It would have been more instructive if he had considered $F(x) = ax$ for $1 > a > 0$ and determined when his theorem applied.

So the situation is clean and crisp but does nothing to suggest a meaningful class for which Theorem 2.4 provides new and useful information. What it did do was spawn a large and largely polluted literature. ◇

3.2 Nonsmooth Examples

Example 3.4 (Critical points)**.** A real-valued differentiable function on the line with a unique critical point which is a local minimum has a global minimum at that point. This is no longer true if the domain is the plane [11, p. 19]. Consider

$$f(x, y) = x^2 + y^2(1 - x)^3.$$

It has a unique critical point in $\mathbf{R}^2$, which is a local minimiser, but has no global minimiser (or maximiser).

Before generalising concepts one is well advised to understand the concept being generalised. ◇

Example 3.5 (The convex subdifferential)**.** The value of convex analysis is still not fully appreciated. It is the rare undergraduate analysis text that includes Theorem 2.6. From this result the entire convex calculus flows and much more [29, 11, 13]. In [1] we try to survey examples less well-known to optimisers.r

Central to the max-formula is the fact that the subdifferential is a global object while the directional derivative is locally defined. A fortiori, every critical point is a global minimum (see Example 3.7.) ◇

Example 3.6 (The Clarke subdifferential)**.** By the time Clarke produced his work, the value of the convex result (Theorem 2.6) was well appreciated. So when he observed that $D_h^c f(x) = D_h f(x)$ for continuous convex functions his justification was already well started. When Clarke observed that a function with continuous first-partials is continuously and so strictly differentiable it meant that Theorem 2.7 also recaptured Cauchy's Theorem 2.5.

The function

$$x \mapsto x \sin\left(\frac{1}{x}\right) \tag{7}$$

defined to be zero at 0 provides an example of a differentiable Lipschitz function with $\partial f(0) = [-1, 1]$ [30, 18, 11, 14]. Variations on this theme

$$x \mapsto x^\alpha \sin^\beta\left(\frac{1}{x^\gamma}\right)$$

for various $\alpha, \beta, \gamma > 0$ are a rich source of limiting examples.

Example (7) nicely illustrates that strict differentiability is central, ex post facto, to Clarke's theory. At this point Simon and Russell would be persuaded that Clarke's new subdifferential concept has legs. ◇

Example 3.7 (Invexity I)**.** Clearly all convex functions are invex, as are all functions with no critical points– grace of Theorem 2.9. But I challenge any reader to supply (a) a useful class of non-convex invex functions or (b) a worthwhile sufficient condition for invexity. I reserve the right to define 'useful' or 'worthwhile' if I am sent any such an example. I promise to be fair. Yet there are published papers on even more elusive topics as *semi prequasi-invexity.*

The study of critical points even for smooth functions is subtle and involves fixed point and variational or mountain pass techniques [14, §4.6 & §6.4]. As the next theorem shows there is no invexity-based free lunch. ◇

Theorem 3.8 (Critical points [14, 10]). *Let S and B be respectively the closed unit sphere and ball of $\mathbf{R}^N$. Then for any continuously differentiable function $g\colon B \to \mathbf{R}$, there exists $w, z \in B$ such that*

$$\max_{a\in B} \frac{g(a) - g(-a)}{2} \geqslant \|g'(z)\|, \tag{8}$$

and

$$\max_{a\in S} |g(a)| \geqslant \|g'(w)\|. \tag{9}$$

There is, however, a smooth function on B with $f(a) = -f(-a)$ for all a in S but with no critical point in B [10]. Theorem 3.8, and much more, remains true for locally Lipschitz functions on replacing g' by $\partial^c g$ [14, §4.6]. However, as noted in Theorem 2.8 the Clarke subdifferential – or even the limiting or approximate subdifferential – generically contains no information at any point other than the local Lipschitz constant [14, Theorem 5.2.23]. So every point of such a function is a critical point for the given subdifferential.

Example 3.9 (Invexity II). Finally let us observe that invex problems have no useful permanence properties. Unlike locally Lipschitz or convex functions, the uniform limit of invex functions need not be invex. Consider $f_\epsilon(x) = x^3 + \epsilon x$ defined on $[0, 1]$. Then for $\epsilon > 0$, the function is invex as it has no critical points, but the uniform limit f_0 is not. ◇

3.3 Computational examples

In this final subsection I wish to make some very brief comments about numerical examples. Much of what concerns me and others is summarised in [3]. Of course generalisations of algorithms should be treated to the same scrutiny as any other generalisations. Are they needed, what are their new features, and so on? But there are other issues specific to illustration of new algorithms.

Many papers with proposed and presumably improved algorithms do not contain detailed tests. Rather they settle for a few often small examples say in MATLAB. Unless the comparison is with well implemented versions of the competition such examples are pretty much worthless. Things to avoid include:

- Iteration counting unless that is a truly robust measure. There are so many ways to count steps. Operation/flop/bit count along with the speed and number of processors is a little more instructive but is still fraught.
- Presenting only the most favourable results not the typical performance or the proportion of failures.
- Giving too little information for an interested researcher, should one exist, to replicate or reproduce your results. At the end of a paper written in the mid-fifties of the last century Newell and Simon remind the reader to switch on the computer.

I used to think the phrase "toy example" insulting, but in a world where we can all run thousands of tests of most implementations, if what you have is a "toy" proof of concept, please say so.

4 Recommendations and Conclusions

> When the facts change, I change my mind. What do you do, sir? – John Maynard Keynes (1883-1946)[4]

Remember that times and circumstances change. So while it is important to have critical views like Keynes we should be able to admit when we were wrong. There are topics that I once found deathly dull that now thrill me. There are new ideas and changing environments, say computationally, that make old methods once again relevant.

So to finish here are a few recommendations or opinions.

(a) Not all questions deserve to be answered.

- Early in my career I submitted a short paper comprising a cute construction in Banach space geometry. It answered a published question posed by Dr A. Somebody. The paper was quickly rejected with the report "Not all questions deserve to be answered." This was mortifying and perhaps a tad unfair but in the long term that rejection did me a powerful lot of good.

(b) Aim to have two qualitatively different examples. Ask if they are natural or contrived.

- Thirty years ago I was the external examiner for a PhD thesis on Pareto optimization by a student in a well-known Business school. It studied infinite dimensional Banach space partial orders with five properties that allowed most finite-dimensional results to be extended. This surprised me and two days later I had proven that those five properties forced the space to have a norm compact unit ball – and so to be finite-dimensional. This discovery gave me an even bigger headache as one chapter was devoted to an infinite dimensional model in portfolio management.

  The seeming impass took me longer to disentangle. The error was in the first sentence which started "Clearly the infimum is ...". So many errors are buried in "clearly, obviously" or "it is easy to see". Many years ago my then colleague Juan Schäffer told me "if it really is easy to see, it is easy to give the reason." If a routine but not immediate calculation is needed then provide an outline. Authors tend to labour the points they personally had difficulty with; these are often neither the same nor the only places where the reader needs detail!

  My written report started "There are no objects such as are studied in this thesis." Failure to find a second, even contrived example, might have avoided what was a truly embarrassing thesis defence.

- *Do all objects with P have property Q* questions perhaps only need one negative example but as described in Imre Lakatos's classic 1976 book *Proofs and Refutations* [22] in mathematics there is a healthy and necessary interplay between theory building and example production ('monster barring').

(c) Better an interesting new proof [20] of a substantial known result than a modest and routine generalisation of an uninteresting result.

[4] Quoted in "Keynes, the Man", The *Economist*, December 18 1996, page 47.

- A.E. Young commented that a second proof of a known result is always welcome; in part because the first is so often wrong.
- If the generalisation is not compelling does the proof introduce novelty or needed complexity?

(d) Don't imagine many people are reading your paper linearly. Most readers – if one is lucky enough to have any – are leafing through looking for the punchlines. So avoid too many running hypotheses or at least make a full statement of each major result.

- And check for such running hypotheses. For instance, Rudin [31] assumes all topological vector spaces are separated while most authors do not. This can lead to careless misapplication of his theorems.

(e) Remember that your readers or audience may well be using English as a second language. This does not mean you should dumb down your language but – as with the advice to restate your main hypotheses – the key points should be made in simple declarative English.

- Even for native English speakers, this is prudent. For example, in referees' reports one frequently reads that the work is "quite interesting" or "quite reasonable". This is problematic, since in American English "quite" means "very" but in British English it is more likely to mean "not very". If used by a British mathematician who has worked for a long time in North America, who knows? So if you mean "very" write "very".

(f) Most research mathematicians are not scholars let alone trained mathematical historians. Avoid the temptation to say that an idea was invented or introduced by someone – whether it is Hilbert or your supervisor. Say rather that you first learned about the topic from a paper you cite.

- In 1981 I had to try to trace the history of various ideas and errors in the early history of semi-infinite programming. It was only then twenty to twenty five years old. I spoke directly to most of the protagonists and nobody was really sure. So it is most unlikely that a researcher today can really disentangle a piece of 50 year old mathematics history. Better to be safe and ambiguous than to be certain and wrong. Euler does not need more concepts named after him and Borwein probably does not deserve it!

(g) Make sure your citation list is up to date. In the current digital environment there is no excuse for failing to do a significant literature search.

- If it comprises largely papers by your clique explain why. Similarly if all the citations are decades old, explain why – if you can. Otherwise, the editors will not be impressed.
- A list of papers saying *earlier related work is to be found in [13, 11, 26, 30]* does not make an adequate literature review. And using the word 'applied' does not by itself make it an applied paper.

(h) When you submit your well-motivated and carefully written paper (including a reasonable literature discussion and great examples) to a journal remember that you alone, and not the referee, are responsible for correctness of your arguments.

- The editor typically asks for guidance about the article's suitability for publication in the given journal. The more interesting the referee finds your paper, the more likely he or she is is to check many of the details.
- If you estimate how long it takes you to read a substantial paper from end to end, you can see that as a referee you simply do not have the time to perform a 'forensic' refereeing job more than once or twice a year at most — unless you are willing to make refereeing your primary intellectual task.
- Indeed, while our literature is perhaps the most robust of scientific literatures it is not entirely reliable and one should read each paper with 'attitude'. If you admire the authors, you read to see what the new ideas are. If you do not trust the authors, you may look for the holes that explain why they can do something fairly easily that you had failed to do, and so on.

(i) Above all be honest.

- If you do not know something say so and try to explain the obstruction.
- If you have a plausibility argument do not dress it up as a rigorous proof.
- If you have dabbled with computation do not oversell that part of your work.
- I leave undiscussed the many statistical sins being daily committed [34].

Acknowledgements The author wishes to thank Brailey Sims and Joydeep Dutta for many stimulating discussions during preparation of this essay.

References

[1] F. Aragon, J.M. Borwein, V. Martin-Marquez and L. Yao, "Applications of Convex Analysis within Mathematics." Special volume in honour of J.-J. Moreau. *Mathematical Programming B* **148,158** (2014), 49–88.

[2] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives." *Pacific J. Math.* **16** (1966), 1–3.

[3] D. Bailey, J. Borwein, and V. Stodden. "Facilitating Reproducibility in Scientific Computing: Principles and Practice." Chapter in *Reproducibility - Principles, Problems, Practices.* H. Almanspacher and S. Maasen editors. John Wiley & Sons, New York (2015).

[4] S. Banach, "Slatur les opérations dans les ensembles abstraits et leur application aux quations intégrales." *Fund. Math.* **3** (1922), 133–181.

[5] M. Barnsley, *Fractals Everywhere*, Academic Press (1988).

[6] H.H. Bauschke, and P.L. Combettes,*Convex analysis and monotone operator theory in Hilbert spaces.* CMS/Springer-Verlag, New York, 2011.

[7] A. Ben Tal and B. Mond, "What is invexity? Bulletin of the Austral. Math. Soc. Series B. **28** (1986), 1–9.

[8] E. Bishop and R.R. Phelps, "A proof that every Banach space is subreflexive." *Bull. Amer. Math. Soc.* **67** (1961), 97–98.

[9] J.M. Borwein and K. Devlin, *The Computer as Crucible: an Introduction to Experimental Mathematics*, AK Peters (2008).

[10] J.M. Borwein, I. Kortezov and H. Wiersma, "A C^1 function even on the sphere with no critical point in the ball," *International J. Nonlinear and Convex Analysis*, **3** (2002), 1–16.

[11] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization : Theory and Examples.* Springer, (2000) (2nd Edition, 2006).

[12] J. M. Borwein, W. B. Moors and Xianfu Wang, "Generalized subdifferentials: a Baire categorical approach." Transactions AMS, 353 (2001), 3875–3893.

[13] J.M. Borwein and J. D. Vanderwerff, *Convex Functions : Constructions, Characterizations and Counterexamples.* Cambridge University Press, (2010).

[14] J.M. Borwein and Qiji Zhu, *Techniques of Variational Analysis,* CMS/Springer-Verlag, 2005. Paperback, 2010.

[15] R.M. Brooks and K. Schmitt, "The contraction mapping principle and some applications." *Electronic Journal of Differential Equations*, Monograph **09**, 2009, (90 pages). Available at http://ejde.math.txstate.edu.

[16] R.C. Brown, *Are Science and Mathematics Socially Constructed?* World Scientific, 2009, p 239.

[17] W.A. Kirk and B. Sims. *Handbook of metric fixed point theory.* Kluwer Academic Publishers, Dordrecht (2001).

[18] F.H. Clarke, *Optimization and Nonsmooth Analysis*, CMS/Wiley-Interscience (1983).

[19] G.H. Hardy, *A Mathematician's Apology*, Cambridge (1941).

[20] J. Jachymski, "Another proof of the Browder-Göhde-Kirk theorem via ordering argument." Bull. Austral. Math. Soc. **65** (2002), 105–107.

[21] R. Kannan, "Some results on fixed points. II." *Amer. Math. Monthly* **76** (1969) 405–408.

[22] I. Lakatos, *Proofs and Refutations*, Cambridge University Press (1976).

[23] J.E. Martínez-Legaz, "What is invexity with respect to the same η?" *Taiwanese J. Mathematics,* **13**(2 B) (2009), 753-755.

[24] S. B. Nadler, "Multi-valued contraction mappings." *Pacific J. Math.* **30** (1969), 475–488.

[25] W. de Oliveira, C. Sagastizábal, and C. Lemaréchal, "Convex proximal bundle methods in depth: a unified analysis for inexact oracles." *Math. Program.* 1**48** (2014), no. 1-2, Ser. B, 241–277.

[26] J.-P. Penot, *Calculus without Derivatives*, Springer-Verlag (2014).

[27] B.E. Rhoades, "A comparison of various definitions of contractive mappings." *Trans. Amer. Math. Soc.* **226** (1977), 257–290.

[28] B.E. Rhoades, "Contractive definitions revisited. Topological methods in nonlinear functional analysis." (Toronto, Ont., 1982), 189–205, *Contemp. Math.*, **21**, Amer. Math. Soc., Providence, RI (1983).

[29] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N. J. (1970).

[30] R.T. Rockafellar and R.J.B. Wets, *Variational Analysis*, Vol. 317 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag (1998).

[31] W. Rudin, *Functional Analysis*, McGraw-Hill, 1973.

[32] H. A. Simon, *The Sciences of the Artificial.* MIT Press (1996).

[33] C. Zălinescu, "A critical view on invexity," *J. Optim. Theory Appl.* **162** (2014), 695–704.

[34] S.T. Ziliak and D.N. McCloskey *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives (Economics, Cognition, and Society)*, University of Michigan Press (2008).

Jonathan M. Borwein, FRSC, FAA, FAMS
CARMA, UNIVERSITY OF NEWCASTLE, AUSTRALIA
jonathan.borwein@newcastle.edu.au