

A Personal View on Reviewing

KAI-UWE CARSTENSEN

2017

Homepage:

<http://www.kai-uwe-carstensen.de>

URL of this PDF:

<http://www.kai-uwe-carstensen.de/Publikationen/APVoR.pdf>

Email:

contact at kai-uwe-carstensen.de

last change: September 19, 2017

Contents

| | |
|--|----|
| Foreword: What's up, doc? | 4 |
| 1 O journal, where art thou? | 6 |
| 2 Tales of the Unexpected: Breaking bad | 7 |
| 3 How to get away with a million dollar (project-) baby murder | 9 |
| 4 Parad(ise/igm) lost: The butterfly effect | 10 |
| 5 The silence of the lambs: Batman returns | 11 |
| 6 And now for something completely different | 12 |
| 7 The good, the bad & the ugly (the Walking Dead) | 14 |
| 8 Seven samurai: Fair game | 15 |
| 9 The unexpected ignorance of virtue | 15 |
| 10 Pride and prejudice | 17 |
| 11 The edge of tomorrow: The time machine | 18 |
| 12 The waiting . . . : the constant gardener | 18 |
| 13 The judge | 19 |
| 14 Close encounters with the third (kind) reviewer (a most wanted man): Pöppel's eleven | 23 |
| 15 How it should always have ended | 24 |
| 16 Pathfinder: A beautiful mind | 25 |
| 17 Conspiracy theory: Proof | 26 |
| 18 F for Fake: Collateral damage of idiocracy | 27 |
| 19 Trust and Shame | 27 |
| 20 Blind date | 27 |
| 21 Cinderella, Judge Dredd, Inception | 28 |
| 22 Anonymous: Perspectives | 29 |
| Afterword: End of watch | 30 |
| Cold case: A History of Violence | 32 |

Foreword: What's up, doc?

Recently, I read the autobiography of Oliver Sacks. I've always liked his honest, clear, case-study style of writing, and although I lack the corresponding competence in probably all relevant respects (last but not least because of English not being my mother tongue), I will try to do what he did all his life: simply write down what one is concerned with or interested in, just starting at the beginning and then seeing where it leads (although I will not write THAT much, of course).

At the moment, and therefore for this whatever-it-is (an essay with small, strangely-titled sections (?)), this is about reviews and reviewing. I like to read non-scientific reviews, as you can find the best and worst of man's intellect in a few lines of text. With growing scientific experience, I more and more wondered why scientific reviews –against expectations– often tend to the latter.

Most of the following will concern the scientific reviewing of journal articles. One can compare submitting to a journal with going to a doctor. With luck, everything works fine and you are treated well, to everybody's satisfaction. More often than not, however, once you enter the doctor's office, you turn into a 'you know nothing Jon Snow'-nobody who is treated like a dotard, and who has to elicit the doctor's sense of urgency for a treatment (instead of being cured on account of the (obvious) symptoms). In both cases, either you are satisfied, or you have to go elsewhere. Yet in reviewing, you cannot even complain.

If you happen to get the same proofreading services spam as I do, you might know the dramatic description of that situation and its ubiquity (here's a passage of an email I got today, cited without reference, of course):

<drama>

[t]hen the email message you had been awaiting for months arrived from the editor of the top-tier journal to which you submitted your latest paper, and his or her response to your work was not at all what you expected. Instead of commending you for writing such an excellent article and informing you that it would be published in the very next issue of the periodical; instead of even a less positive but nonetheless workable request to revise certain aspects of the paper so that it could be accepted or at least seriously considered for publication, he or she simply rejected your submission. It was a moment of personal and professional deflation, one in which you began to think that the time and energy you invested in the article might have been wasted, and perhaps one in which you felt utterly alone, as though you were clinging with breaking nails to the bottom of the academic or scientific totem pole.

First and foremost in such a challenging situation is the need to recognise the consoling reality that you are definitely not alone. Rejection from publishers is a far more common experience among authors than acceptance is, and each one of those authors, no matter how confident and experienced, feels overwhelmed to some degree by a wave of disappointment when a rejection letter arrives.

</drama>

Getting serious again, I myself have been a reviewer oftentimes. Besides that, almost all of my articles have been finally accepted, so there is no categorical cause for resentful behavior. Accordingly, although I am mostly involved in what I am going to write about, this is not (or at least only remotely) about anger and lamenting, but about curiosity (how can this be?) and reporting. I must admit, however, that the difficulties I encountered with the publication of interdisciplinary papers in general, and with one of my recent papers in particular, were the inspiration for writing about reviews at all.

I am aware of the fact that talking about reviewing is a communicative taboo area. You will rarely hear something like "Wow, got some great reviews of my paper yesterday!", as it is scientifically uninteresting and rather leaves a bad aftertaste. Neither will you hear lamentations like "Oh dear, my paper was rejected for the fifth time!",¹ for obvious reasons. I will have to break this taboo, of course, and you will find honest examples for both positive and negative reviewer comments. Please forgive my trying to strike a balance between the different kinds of impressions this may evoke.

I am also fully aware of the fact that others probably have much more to tell about this scientific everyday phenomenon.² But then, why should I care?

¹Except for ironic reports of negative review(expression)s that can be found on <https://twitter.com/YourPaperSucks> (thanks to Elke Hentschel for pointing me to this thread).

²... and even write papers about it, for example, G. Cormode (2008), How not to review a paper: The tools and techniques of the adversarial reviewer. *SIGMOD Record*, 37(4):100-104; Mark Allman (2008), Thoughts on Reviewing. *ACM Computer Communication Review*, Editorial Contribution, 38(2), April 2008.

1 O journal, where art thou?

Everything we're going to talk about starts with the selection of a high-quality journal for publication. This is probably easy if the topic of your article perfectly fits some journal's aims and scope. It gets more difficult if you're spoilt for choice between more of them. It gets tricky if you don't find a journal (at once) and have to adapt, either in the presentation of your article to some journal's editors or even in your article's content, or by making concessions to the desired journal's quality or scope. And it gets nasty if you really don't fit, for example because you are working interdisciplinarily (partial fitting at best) or your stuff is new, not matching the basic assumptions of any journal at all. I am especially interested in these two latter options.

For decades, I have been working in the interdisciplinary field of cognitive science. To give an impression of what that is, see the text from a leaflet I designed some years ago (actually, as of the time of writing this, the text is more than 17 years old and still used –only slightly modified– for advertising the corresponding course of studies at the University of Osnabrück):

More than two millennia ago, philosophers began to think about the properties of mind and matter.

A few centuries ago, the idea to view thinking as mental calculation arose.

With the advent of computers in this century, this idea could be realised.

In recent decades, members of various disciplines have collaborated in studying mental processes and representations and in trying to build intelligent systems.

In the last few years, it has become obvious that it is also necessary to study the brain and its neural architectures.

Now, Cognitive Science has emerged as an 'inter-discipline' concerned with the scientific study of mind and brain, the explanation of human linguistic and non-linguistic behavior, and the building of artificial intelligent systems. It comprises central aspects of the following disciplines: anthropology, artificial intelligence, (neuro-)biology, computer science, (computational) linguistics, mathematics, (computational) neuroscience, philosophy, (cognitive) psychology.

Working in Cognitive Science means more than just working in one or two of the participating disciplines. Ideally, someone working in this field has basic, multidisciplinary knowledge about the fundamentals (problems, goal, methods) of *most* of

them. It also means that one is investigating a topic crossing the disciplines whose problems can be better solved from an interdisciplinary perspective.

As to journals, they mostly reduce this complexity by setting a narrower disciplinary or thematic focus. You're fully right if you believe that this may lead to problems in publishing an article. Here's an example: I first submitted my paper about *Cognitivist ontologies* to the journal *Cognition* and was rejected because of the following reason given by the editor:

Although the work does afford theoretical constraint, the work is not sufficiently broad to be accessible to the general readership - I believe that the audience at *Cognition* is not the right one for this work, and that this should be submitted to a journal that has closer links to information science, philosophy, or theoretical linguistics.

Then I submitted it to the quite specific journal *Applied ontology* (which, if you happen to know it, is in the intersection area of information science, philosophy, and theoretical linguistics) where it was rejected for many specific, but misleading, reasons. It was finally well accepted in *Cognitive Processing* (!).

2 Tales of the Unexpected: Breaking bad

I don't remember my first scientific review. I only remember my first *important* scientific review. It was definitely a good one, but it had bad consequences. That is, I tried to be objective (though not necessarily polite) by not only emphasizing the bad aspects of the reviewed paper, and elaborate by commenting all identified mistakes and not just bashing the paper on the basis of some. This turned out to be a mistake, and I would certainly write it differently today, given the experience I have now.

There's a background story that starts with a talk I gave at a European Artificial Intelligence conference where I presented a *précis* of my master's thesis. I met young colleagues who invited me to their institute's colloquium for an extended presentation. A seemingly nice guy, let's call him Walter Matthau, who had just gotten his informatics diploma, asked for a copy of my thesis, which I happily sent to him. He was "raised" in a top-notch research group, so his interest rather flattered me. And as I said, he looked like a good guy.

At that time, I had become a member of the reviewing board of a thematically related conference. Guess what happened. A few months later I got a paper to review, written by this very guy. As can be expected, the structure and the style of the text was highly professional. The content, however, was an unbelievable

impertinence. For the most part, he presented information from my thesis and cited the corresponding sources (mixed with others). Awkward passages, strange misuse of terms, misrepresentations of content, and typos (e.g., of journal names) showed, however, that he had just copied without understanding (this being an instance of plagiarism). Furthermore, although the additional, “unique” aspect of his paper was an eye-catcher, his theoretical proposal was not more than a loudmouthed promise of things he wanted to develop.

There was an accordant rating scheme (structure, content), and I gave an A and an F, correspondingly. I should either have been dishonest (downgrading the structure part) or else explicitly disqualify the paper. Instead, with me the only expert for judging the faults and with the other reviewer(s) probably impressed, the paper was accepted. Sometimes it is indeed appropriate not to mention or highlight positive aspects of a paper.

There is another great disadvantage of such benevolent “good” reviewing, which I became aware of only afterwards. In listing all faults and (inferred) corrections in my comments on the paper, I helped to iron out its weak parts and tremendously improved its quality. As a reviewer, you may not be thanked for that (I, for one, wasn't). And if you are a competitor, this may just be a silly thing to do.

This small mistake in reviewing had a tremendous effect on my career: Walter became the new expert of the field, and although I had moved to other research topics anyway, my just beginning “fame” in that field was swept away. He had, by the way, managed to only mention me in a footnote, only with regard to a minor aspect, and only derogatively. But it was somehow fun³ to see (and remark) that he had cited a totally irrelevant paper (the author happened to have my surname). Here's the corresponding excerpt of my review (the full annotated text of the review is [here](#)):

- I don't know whether H. Carstensen with his paper on 'The Complexity of Testing the Equivalence of Transition Sequence' is relevant for this topic !?!
- I don't know the journal 'Environment and Behavioral' (->Kuipers)
- I don't know the journal 'Environmental and Behaviour' (->Leiser/Zilbershatz)

Suffice it to say, after some more loudmouthed papers, Walter grew quieter, eventually presented his dissertation thesis (scarcely anyone talks about), left the field and later got a professorship in an entirely different discipline.

My bad.⁴

³Isn't it ironic?

⁴A colleague tells me that even worse once happened to him: the paper he had submitted to a conference was rejected with strange justifications, yet to his astonishment and annoyance, he

3 How to get away with a million dollar (project-) baby murder

There was a time when one project I had worked in ended and the other (an expected follow-up) didn't come. We had connections to groups in other disciplines and eventually came up with the idea of applying for some *real* money, that is, to get a grant for a big interdisciplinary project (roughly, a million dollar amount). As I had not much to do, I was the one who gathered all information and wrote the proposal (and no, the picture of the bird wasn't my idea).

The topic was scientifically located in my old research field, so writing somehow felt like coming home. I managed to find a theme for the project that not only embraced the expertise and current work of the participating groups, but also presented a new and interesting research perspective. In a nutshell, the goal was to build a system that would produce multi-modal route descriptions of different granularity adapted to the requirements of a user in a certain situation (as opposed to rigid, context-free, off-the-shelf standard texts). Combined with our intention to emphasize the possible synergy of our collaborative work, this seemed to match the grandeur the application was expected to show.

On the other hand, it was somehow clear that this was more a let's-try-to-get-some-money patchwork thing than a product of a grown collaboration, and that this might shine through. With only scarce experience in writing project proposals (let alone of that size), I was unsure and quite pessimistic about whether I had found the right level and tone between textual bravado (see, one *has* to show off a bit to compete with fellow-applicants) and self-fulfilling prophecy of rejection. So when we finally found ourselves devoid of acceptance, disappointment should have been kept within a limit.

Beforehand we had been allowed to supply a list of possible reviewers. Our first reviewer (which I recognized easily) actually was one I had warned the others to put on the list. He was highly competent and respectable, but as a psychologist, his level of interest in fine-grained testable assumptions was simply incompatible with our multidisciplinary research methodologies in this big project. And so it came. His review was decent, highly professional, objective (weighing pros and cons) and said what could be expected: that our proposal was very interesting, cutting-edge research, but a bit too ambitious and complex. As could have been expected, too much bravado for him.

found a Walter-Matthau-style adaption of his text in the proceedings of the same (!) conference, authored by another person.

But I definitely did not see the second reviewer coming. I am not referring to most of his little one-and-a-half-page text where he complained about being given too little information about details, which was named as one reason for rejection (you must know that the proposal had a maximal length of 20 pages that we fully exploited . . .), or to his non-interdisciplinary 'think small' viewpoint. I mean his text as being an instance of a more general review phenomenon.

What does a reviewer do if he has not much competence and/or not many good arguments for a rejection (which he has categorically chosen for some reason)? He tries to emphasize faults (although they are small and less relevant) in order to make the paper/proposal look bad! And that is what he did. The annoying part, however, is that it was unjustified to do so, hence, that he was wrong! The following is from a frustrated e-mail I wrote to my colleagues afterwards:

one detail is especially important: we are accused of citing a source that allegedly gives incorrect information ('the quote on p. 4 from Höök is incorrect') and are given references we should have read. **HOWEVER, THE REFERENCES ARE FROM 85,86, AND HÖÖK IS FROM 91 AND HAS ALL OF THE MENTIONED REFERENCES IN HER BIBLIOGRAPHY!!!**

Again, there is a fun part (see Fn. 1) of this story. We had to decide whether to retract the proposal or not within a week. Unfortunately, the results were put in a dead letter box (!), so we/I didn't get the reviews in time. We therefore rather retracted the proposal.

So we had planned a natural language (!) processing (!) system development (!) project and were rejected by a psychologist and a geoscientist (whatever exactly) who very likely had little experience with natural language processing system development projects. Although we presumably wouldn't have vetoed the second review anyway, this left a bitter aftertaste for me.

With respect to reviews in general, this points to the importance of the reviewers' competence in and attitude towards a certain research field. Every intelligent person is able to produce a reasonable opinion piece on request. But would you really ask a vegan for his opinion on the taste of a certain Wiener Schnitzel made of veal?

4 Parad(ise/igm) lost: The butterfly effect

Science is in various respects like religion. The god prayed to is analyticity, and the religions, which vary in space and time, are the scientific paradigms.

In a paradigm, some basic beliefs need not be questioned. On the one hand, this has many advantages, as intellectual effort can be freely and productively brought to bear within this range. On the other hand, however, these beliefs *must* not be questioned, and any violence of this rule corresponds to a heresy that is to be punished.

Like a religion (or even, a sect), a paradigm provides a home to its adherents, and they behave correspondingly: other members of the (peer) group are treated friendly as guests, strangers are treated with suspect or, if identifiable as outsiders, with intolerance, refusal or even hostility.

All this is as natural as the usual social behavior in the kindergarten. What is often overlooked, however, is that both science and religion undergo changes and developments, which are no less “natural” (a fact that adherents to some paradigm will probably be reluctant to admit). From this global point of view, it only takes a bit of consideration to notice that a lot of injustice is involved here: without proponents of different ideas there will be no change, but those who propose a change are heretics in the current paradigm. They may be vindicated sometime later, but for most of them this never happens (or perhaps too late). In such a situation, one can only hope to be the butterfly heretic affecting the paradigm weather who becomes the one cited/famous for apparently having caused the change.

As to reviews, which somehow exemplify this global scientific phenomenon, an author may be affected in this very respect. He/she may not share the basic beliefs, or may present results that contradict them, or he/she may only be suspected of either one of these conditions. Many papers are probably not rejected because of their objective (lack of) quality but merely for their inaptitude to the current paradigm.

It is certainly hard to tell ignorance from reluctance, but with some (bad) luck one can sometimes virtually smell the fear of editors and/or reviewers of losing their paradigm (their paradise) as it is. No further proof needed for why this in most cases ends badly for the author (although fortunately not at the stake anymore).

5 The silence of the lambs: Batman returns

Let's talk about traumata. In real life, a politician might be bashed because he gave a frustrated, but honest interview while his lying, smiling opponent wins the day. A singer might be booed because he/she didn't strike the right note in spite of having given a good concert overall. Somehow, life is not fair in general, and there are many reasons why someone dies of stomach cancer or becomes drug-addicted, a superhero or whatsoever.

In science, every unjustifiedly⁵ negative review can be a traumatic experience. The best way to handle this (probably practiced by most) is the “whatever” method: simply ignore and go on.

I have opted for a slightly different resort, for the following reason. In case of acceptance, an author is asked to consider the comments of the reviewers, revise the article accordingly, and give a report of what was changed, e.g.:

If you can submit a new manuscript to us, please provide us with a careful discussion of how you have addressed the comments [...]

No converse procedure (assessment of reviewers’ comments) is envisaged in principle.⁶ I simply use to ignore this fact and send back annotated comments, asking the editors to send them to the reviewers (which is probably never done).

I know this is like asking for a doggy bag after a meal in a star-rated restaurant, and the editors use to sniff at that or complain, and of course (politely say that they) refrain from doing so. But it sends a clear signal about the editors’ choice or the competence and thoroughness of the reviewers that might lead to an overall improvement. And it keeps me in control and sane, of course.

Once I was so bold to ask the editors to reconsider (I do not do this anymore) because of the sheer mass of ‘mis’es in the reviews (*mistake, misunderstanding, misinterpretation*). I got a polite reply on my re-comments that tried to give a justification of the reviewing process as it is, and in which it was stated that reconsideration was inconceivable. I found all that unconvincing. This is what I (imperfectly) wrote back:

From an author’s point of view, I would then kindly ask you to instruct your reviewers to not come up with statements (in the first place) whose necessary revision afterwards has no impact.

So, if I end up as a scientific lamb getting butchered, hear me scream, and watch out!

6 And now for something completely different

Is there a difference between reviewing *student* papers and reviewing *journal* articles? In my opinion, there isn’t. It’s all about competence and the quality of comments, everything else being entailed. In both cases, therefore, authors should receive high-quality comments on the paper covering all relevant aspects.

⁵ I am not talking about the trivial case of bad papers(’ reviews) here, of course.

⁶with the exception of conferences who collect authors’ ratings of reviewers for granting a *Best Reviewer Award*.

In reality, there is a big difference, however, as modern peer reviewing for journals unfortunately has some negative aspects. First, the reviewing process nowadays is mostly automatized via some Web-editorial manager. From the beginning, this puts up a barrier between author and editor and hinders loyal communication. Second, usual reviewing is double-blind. Originally, this was probably meant to serve objectivity. Effectively, however, it primarily leads to opaqueness: Editors cannot be called to account for their choice of reviewers, and reviewers do not need to pay much attention to the quality of their texts. This not infrequently results in textual berserk behavior, because nobody cares (except the author of the reviewed article). I simply hope that this does not happen in my/your communication with students.

Interestingly, some journals return to non-blind reviewing in these years. Correspondingly, author and reviewers might interact openly, the reviews might be posted on the web, and/or the reviewers might be named in the published article. As a result, only people feeling competent in a certain field should accept an invitation to review, and reviewers will certainly watch the quality and the style of their comments (more closely). Based on my experience, I can definitely recommend this procedure.

Obviously, I won't talk much about *conference* paper reviews here. It is well-known that every Tom, Dick, and Harry may get selected for this task, and I have evidence that this reduces review quality substantially. When I tried to present the central results of my dissertation thesis to some pertinent interdisciplinary conference, I was rejected, the reaction roughly being "what's that/new?", "don't understand", "not good enough" or the like. For the record, the same stuff has been published later in journals, even though only when the time was more ripe for it. And yes, that made me very angry.

By the way, I do not claim to have always given good conference paper reviews. They are sometimes a pain in the proverbial... But I have proof that I did so at least once or twice ;-):⁷

Lieber Herr Carstensen.

Vielen Dank für Ihre präzisen Gutachten der COLING-Proposals, sie sind sehr wichtig fuer die Entscheidung. Ich wuerde mich freuen, wenn die anderen Gutachter auch nur annähernd so prompt wären wie Sie.

Mit besten Gruessen,

Helmut Schnelle

While we're at it, there are also *book* reviews, which constitute an entirely different sort of text. I once reviewed a book (a festschrift, to be precise) about the perspec-

⁷ An email from 1998 roughly saying 'well done!'. You might be interested in [the vita](#) of Helmut Schnelle.

tives of computational linguistics containing more than 50 small opinion pieces. Quite unreviewable, one could think. Here's the reaction of one of the authors of the book:⁸

Lieber Herr Carstensen,
 ich hätte nie gedacht, dass man dieses Buch ernsthaft und konsistent rezensieren könnte. Das ist Ihnen in bewunderungswürdiger Weise gelungen: informationsreich, kohärent, kritisch, weiterführend und gut lesbar. [. . .]
 Mit freundlichen Grüßen
 Ulrich Schmitz

Interestingly, my last published text (at the time of writing this) is a 6000+ words (!) long review of an introductory textbook (preprint).

7 The good, the bad & the ugly (the Walking Dead)

I have found out that there are three kinds of reviews: the good ones which make you happy, the bad ones which make you sad, and the very bad ones which make you angry.

Of course, those leading to acceptance almost always belong to the good ones. Interestingly, negative reviews may sometimes be categorized as 'good', too. In that case, I am happy about the competent feedback I get, the effort that is spent or the empathy that can be felt (despite the article's shortcomings).⁹

There is not much to say about the bad ones. I get sad either because of me (unhappy) or because of the reviewer (dismayed)(inclusive *or*, of course). As to the latter, there's often these 'mis'es again (plus *mischief*, *misery*, and miscellaneous). As far as I am concerned, bad reviews are neither fish nor fowl, and the sadness appears to me as a rather boring feeling, corresponding to being in a limbo. Whatever that exactly is, I suppose it has the same 'you cannot do anything about it'-air.

Reviews of the third category are definitely different. I love to hate them. Compared to the dull limbo, this is like being in hell, cracking the skull of every zombie that comes running to you (no, I'm not a 'Walking Dead' fan). Speaking of zombies, very bad reviews can be *that* ugly. They may either have deficiencies in what should be there (virtues) or show abundance in what should not be there (vices).

⁸ In a nutshell: 'Wow, you did it, and quite good so!'

⁹The converse is true for positive, but incompetent reviews.

8 Seven samurai: Fair game

It is not so easy to get a grip on what the virtues are that a good reviewer should possess, apart from what one should always expect (intelligence, competence, skill). After looking for a while I found the seven virtues of the Bushido code, six of them being relevant here (with *honesty* being a too general virtue).

Rectitude/Integrity simply means 'always do the right thing'. For a reviewer, this amounts to giving an objective estimation of the author's article and a thorough and clear justification of that estimation, all based on competence and close reading.

Courage is needed when there is a conflict between the reviewer's knowledge and the article's results, with the latter being correct. The courage then consists in admitting this fact instead of obfuscating it by focusing on irrelevant aspects leading to rejection.

Benevolence can tip the scales if there is no conflict, but the author adheres to a different theory, paradigm or the like. Reviewers can make authors look bad, but should not do so.

Respect/politeness should be a prerequisite for all social interaction. It is easy to ignore this basic aspect in double-blind procedures, though.

Honor becomes relevant in the reviewing function per se. A reviewer has earned it and must (according to the code) enjoy it. A reviewer who is reluctant to fulfill the function wholeheartedly and earnestly does not show this virtue. In that case, this person might forget the important role she has in the scientific system and might not act adequately.

Loyalty is relevant in the sense of giving the author as a colleague an unconditional chance in the process. Here, it may lead to conflicts, though, as the reviewer must also be loyal to his journal. In that case, this may not be beneficial for the author: an article could be rejected not because of its quality but because the reviewer thinks (perhaps wrongly) that it does not fit the journal's aims, scope or addressed scientific group.

To be honest, these virtues seem a bit redundant to me. In our context, one simply expects reviewers to be *fair* in the game of thrones publications.

9 The unexpected ignorance of virtue

Unfortunately, reviewers often ignore the mentioned virtues and rather tend to the other end of the scale of measured morality. Correspondingly, one can approach the discussion of review quality from the question of which 'sin' is committed in each

case. It is interesting that even in the 2900+ year old biblical sins below, there are already a few applicable ones.

Proverbs 6:16-19 King James Version (KJV)

16 These six things doth the Lord hate: yea, seven are an abomination unto him:

17 A proud look, a lying tongue, and hands that shed innocent blood,

18 An heart that deviseth wicked imaginations, feet that be swift in running to mischief,

19 A false witness that speaketh lies, and he that soweth discord among brethren.

A *proud look* corresponds to the pride (or hubris) a reviewer may show if he (always) regards his competence as superior to the author's one. One has to keep in mind that the author can also be expected to be competent. Compare this situation to a speed car race. Here you do not get the pole position just by having won a race sometime before. You have to earn it anew each time! Accordingly, a reviewer should always check whether he really has this position or whether he is one of those behind. There is no maximum-competence guarantee just by being a reviewer. There is, however, some chance of becoming an arrogant prick.

Devising wicked imaginations is more or less a paraphrase of misreading or misinterpreting the article. The resulting allegations are disastrous, because once on the road of misunderstanding, the reviewer will quite probably never say something like 'in spite of that I recommend acceptance', and there is no means to object.

The fifth and sixth ones are the equivalence of judging/rejecting too quickly and/or without reason, i.e., despite author's correct results and argumentation. The final sin points to the inevitable social consequences bad reviews produce: in small-scale scope, this is the frustration of the author (remember, we are talking about reviews that make you angry), in large-scale scope, this may lead to 'paradigm wars' where articles of one paradigm are not published in journals of the other.

Among the *deadly sins* mentioned later in religious history are further vices: there is *Greed* if the reviewer's decision is influenced by thinking about his own advantage. *Sloth* can be found in the omission of looking for, respecting or presenting positive aspects of author's article. *Wrath* underlies the reviewer's flying into a frenzy because of minor faults (under the presumption of wrongness of the article, although its main thrust is in fact correct). *Envy* may be the reason for not supporting an article that presents more elegant, practicable or general solutions than reviewer's own theory.

At the end of the day, it is probably a mixture of ignorance in detail (even despite competence in general) and slothfulness that is at the core of a bad review. Having

some more of the others will make it an ugly one. Rest assured, if all this is paired with small-minded, arrogant nit-picking, it will drive you crazy.

Again, I don't know whether I myself always live up to the expectation of being virtuous, not vicious, in reviewing. See for yourself, [here](#)'s a recent (anonymized) review of mine.

10 Pride and prejudice

Let me add another perspective to the aforementioned points. There is no doubt that *skepticism* is one of the basic principles of scientific procedure in general and of reviewing in particular. It is simply right to be careful in judging new proposals and hesitating in approving them. Unfortunately, there is reason to assume that this beneficial principle may take on its own life and morph into a loathsome trait of some reviewer/editor, for the following reasons.

First, everyone in any profession undergoes some character development: with your work done, your positions filled and your experience gained, there is a natural change in self-esteem. If positive, you will probably get proud of what you have done and become. Nothing to object to that, of course.

Second, while you may have become successful, the number of your tasks and functions will have increased. Accordingly, this reduces the available resources for every single part of your agenda.

Third, with increasing work load and time gone by, you may have missed some developments, as well as the fact that some of your favorite theories may already be outdated.

Now apply all of this to reviewing. I have observed so often that experienced people judge too quickly, because of self-confidence, lack of time, or half-lived competence. As reviewers, such people may therefore be skeptical about a paper they did not understand, yet not for scientific, but personal reasons. Here we have the *proud look* from above, where pride turns skepticism into prejudice.

While the foregoing description evidently applies to the aged researcher, I just realized that a similar mental attitude may also be adopted by young researchers. They probably have dived deep into some topic, have written a complex dissertation thesis, have been asked to give talks and write reviews and so on. Being proud of what they have achieved, however, they may not (yet) realize how much they do NOT know, and therefore act inappropriately. I just hope that I have not been one of that kind 20-30 years ago although I fear I was. Unfortunately, while everything has its costs, arrogance comes for free. You just have to pay for it later.

11 The edge of tomorrow: The time machine

Did you ever ask yourself why doing a top-notch job in science is called *cutting edge research*? A simple answer could be: if you're doing it, you're on the verge of success, but you may also be cut off from opportunities, money, success etc.

It is interesting to observe that reviews more often than not do not reflect this situation. Both editors' and reviewers' behaviors often suggest that the author is sitting comfortably in a streetcar named desire and that he may just take the next one after rejection, here is an example:

I would like to thank you very much for forwarding your manuscript to us for consideration and wish you every success in finding an alternative place of publication.

This does not quite mirror the reality of the author having a hell of a ride on the road to success, with ample chances of missing the boat. After all, the whole writing-and-reviewing *is* being at the edge of tomorrow, as the author (unlike the time traveller in one of my favourite novels/films (guess which one)) hasn't got "all the time in the world". Reviewers may actually be the ones doing the cutting at the cutting-edge in research.

Another important aspect of the cutting-edge metaphor is: it may hurt.

12 The waiting . . . : the constant gardener

Wait, there is another aspect of time in reviewing: you never know how long it takes.¹⁰ While the author is keen to get a quick response, the reviewer will probably put the paper on his/her stack.

However, there is something wrong with the use of the term *stack* here. In information technology terminology, a 'stack' is a last-in first-out storage device, like the push-down coin-storage devices bus drivers use for the small change. If this sense were applicable here, the reviewer might actually grab the paper from the top of the stack, which corresponds to some chance of eventually getting a quick response.

This, regrettably, almost never happens, as far as I know. Usually, reviewers exploit the time of deliverance (around four months), until/unless all older tasks have been accomplished.

We therefore have to view the situation differently: rather than being put on a stack, papers are put on a pile instead, which functions as a *first-in first out* storage

¹⁰and yes, the pun (in the title) is intended.

device (i.e., a 'queue' in IT terms). That is, papers to be reviewed sink down in the pile of to-dos of some reviewer like some rotten banana on a compost heap used in gardening. The question is: will the reviewer treat a paper as irrelevant waste, or as fertile humus for its scientific domain?

Back to the waiting: it can take a long time.¹¹ And just like with restaurants: you are **not** allowed to send your order to two or more providers at the same time and take the first satisfying service, of course.

13 The judge

Now let's talk about the editors. They are responsible for the rough estimation of fitting ("screening"), the determination of the number of reviewers, their choice, and finally, for the ultimate decision based on the assessment of the reviews. I'd never had thought how much can go wrong with just that.

First of all, the judgment of whether the article accords with the aims and scope of the journal. Here, editors may, especially in interdisciplinary contexts, simply not be competent for an adequate assessment of the submitted article. Above I mentioned my *Cognitivist ontologies* article that was rejected for *Cognition* by the editor but accepted for *Cognitive Processing* (trust me, there was no specific focus on processing in that paper, and I knew *Cognition* for 20 years). The editor was a psycholinguist (psycholinguistics typically has a quite narrow focus within cognitive science). I do not blame him for not knowing enough, I blame him for not *realizing* he didn't know enough and for not at least granting me further opinions.

Secondly, there is the number of reviewers. One might guess that for minimum regard to objectivity, at least two reviewers should be assigned. This would mean: a clear decision if both judge either good or bad, no immediate decision if there is a tie between a good and a bad review, and a less clear decision in most other cases (few of them positive). In the second case one would probably opt for a third opinion.

Reality does not seem to conform to this nice model, however. In two cases, I was granted only one reviewer, and both of them judged negatively. So when the articles were rejected, this happened on the basis of one opinion each. *Honi soit qui mal y pense*, you might say: the editor probably added his own, and rightly so. Unfortunately, though, I never saw a second review!

There was another case where I got *no* review (after submitting to *Journal of Semantics*). The editor, let's call him Calvin Klein, at least provided me with an

¹¹Right now, I'm already waiting thirteen months for the decision concerning a submitted paper.

elaborate explanation why he wouldn't consider granting me any.¹² Again, it is in his sovereign territory to do so, in principle. But what if I tell you that this was a resubmission (editors had changed in between), an option that I had been explicitly advised to consider by the previous editors (accordingly, the *Cognitivist gradation* piece got published later)? Here's what they had written:

Rethink and rework the paper, and submit it again at a later stage. You should be aware, though, that if you choose to do so, it will be considered a completely new submission (and would likely be assigned a different handling editor who would probably assign it different reviewers). I would NOT recommend, however, submitting the paper again without taking into account the comments of the current reviewers.

And what if I tell you that the theory of Calvin was explicitly criticized in my article? *Honi soit qui ne mal y pense pas*.

Thirdly, the choice of reviewers. To be sure, this *is* a difficult task, and you probably don't want to be in some editor's shoes. Recently, I submitted an article to *Mind and Machines*, an obviously interdisciplinary Cognitive Science journal. I was happy to see it passing the editor's hurdle and waited curiously and patiently two further months. I had adapted to the journal's aims and scope and was quite optimistic that I had managed the balancing act between the theoretical/abstract and the practical/technical in my paper.

When the rejection came, I therefore was quite disappointed. The reviews were neither good nor ugly, but limbo-bad. There was nothing to learn from or to be angry about, the editor had simply chosen (the) wrong experts. More specifically, he had chosen monodisciplinary experts from the participating disciplines who showed a quite complete lack of understanding of the overall picture. To give an example, here's the critical point made about the part where I motivate the relevance of my research for both theory and application:

The author makes a distinction between theoretical and practical considerations for changing the logic. It's not obvious what the distinction is and it might even be controversial that there is such a distinction. Does the author need to get involved in this dispute?!

In my comments, I wrote "Wow, this is *Mind and Machines* [...] And it's just the motivation part" ...

¹²A colleague tells me that things can be much worse: after eight months, her paper was rejected (without further reason or explanation) simply because the editor found no reviewers!!

The other reviewer also had his coming-out moment. In my article, I criticize a theory as being too complex and point to the aspect of learnability (by 2-3-year olds). His comment goes like this (my emphasis):

Children of 2 - 3 years old, in learning to use quantifiers of various sorts, don't need to learn the semantic clauses for these expressions. Rather, they need to learn patterns of **behavior** that are in accordance with the clauses.

Now you must know that at least classical Cognitive Science is somehow defined by *not* theorizing on the level of "behavior", best exemplified by Noam Chomsky's famous critique (review!) of Skinner's *Verbal Behavior*. The reviewer simply disqualifies himself with his comment.

Finally, there is the decision of the editor. In most cases, the constraints given should make it easy. That is, all bad papers would be rejected, as would all unclear papers, because of the, say, 200 submissions, perhaps only 12 can "survive". The real work probably lies in the choice from the good papers in most cases. As an author, I therefore could understand and accept the message "hey, your paper is good, but there are too many better (fitting)".¹³

Kenneth Church (in an opinion piece about reviewing in Computational Linguistics)¹⁴ has quite nicely put to words that the decision should not simply be based on the impression of non-badness of the paper and the unanimous consent of the reviewers:

Controversial papers are great; boring unobjectionable incremental papers are not. The only bad paper is a paper without an advocate. A paper with a single advocate should trump a paper with lots of seconds, but no advocates. Don't average votes. The key votes are the advocates. Negative votes matter only if they convince the advocates to change their votes.

It is the editor who is responsible for a corresponding procedure in reviewing. Yet there seem to be cases where he should (from an objective point of view) check some review and act correspondingly, but does not.

For example, the same article I just talked about was previously submitted to *Linguistics and Philosophy*. The editor, let's call him T.E.D., not only granted me just one reviewer, he furthermore chose a linguist with an extremely narrow focus. To understand what I mean recall that *Artificial Intelligence* (AI) is part of the interdisciplinary *Cognitive Science* (and my papers are typically written from that

¹³I am aware of the fact that this conception probably does not match the usual reviewing procedure.

¹⁴Kenneth Church (2005), "Reviewing the reviewers", *Computational Linguistics* 31 (4), 575-578.

general point of view), and that it is an autonomous (sub-)discipline of informatics just like nuclear physics is for physics or dermatology is for medicine. Here's what the reviewer wrote about some part of my text:

It is just standard AI—a combination of show-'n-tell, bravado, and caricature of everything in sight, opponents and natural language. [...] Nothing in AI is worth advertisement unless its improvements on its predecessors are cosmic

Do you get the craziness of this assertion? It is a public-opinion prejudice, it is an insult (both to me and to the discipline), it is wrong (I'm not Ray Kurzweil, am I?), it is thoroughly unscientific.¹⁵ As an editor, I would be ashamed of having selected him and would never base my judgment on his opinion. But the editor didn't check, didn't reconsider, not even answered my email. Although I promised him that I would go public with that. So here we are.

You wouldn't expect that this can be topped, do you? Well, when I politely asked the *Mind and Machines* editor to send my re-comments to the reviewers, I got the following answer:

I would do you no favor, and indeed would do you harm, were I to honor your request to pass your comments to the reviewers. That I shall not do.

Let's put it that way: there is a whole scale of possible reactions here. At the upper end, something like 'Oh, seeing your comments, I realize that you must be unhappy about my decision. I am sorry. I will certainly do what you ask me to.' In the middle, simple ignorance. His answer, however, is –despite its appealing diction– definitely at the lower end. It perverts the situation of reviewing my paper: bad reviewer choice, bad reviews, bad decision, but the blame is on me if I object to all that. Here's my reply:

thank you for your answer. But honestly, I do not understand the part 'would do you harm'. As the reviewers were quite polite, I stayed polite, too. And the harm is already done to me.

¹⁵It is also in stark contrast to the detailed and differentiated picture on language-related AI I try to give in my [German introduction to language technology](#).

But of course the reviewer didn't know that. So much for anonymous reviewing.

It also ignores the fact that I have worked on semantics all my life (although interdisciplinarily). That's why I just published a review of one of the world's best introductions to semantics: Carstensen, Kai-Uwe (2017), Rezension von "Löbner, Sebastian (2015): *Semantik: Eine Einführung*. 2., aktualisierte und stark erweiterte Auflage. Berlin/Boston: DeGruyter“, *Linguistische Berichte* 250, pp. 247-262 (above mentioned preprint).

But of course the reviewer wouldn't have known that. So much more for anonymous reviewing.

14 Close encounters with the third (kind) reviewer (a most wanted man): Pöppel's eleven

Sometimes you get more than you think is necessary: a third reviewer. The three situations I recall in which this happened to me turned out to be quite different. The first (conference!) paper was accepted although (as I found out when rummaging around in my files) it was probably a close colleague who gave the worst rating. I know exactly that this was not his field, but he definitely didn't want to put himself in the 'I know this guy and give an OK rating' corner. Phew! Thanks a lot, you (other) guys!

In the second case, the article was finally rejected. It was my only non-blind peer review, so I knew the colleagues and how they voted. Everything was nice despite the rejection (and in spite of my re-commenting their comments), so I did not only tell them when the article was accepted in another journal, I also thanked them in the Acknowledgments. For the one who gave me a positive rating, this was probably some confirmation, too.

The third reviewer in the third case (the first Cognitivist gradation paper submission) belongs in the 'ugly' category. I must admit that the article was not yet in best shape at that time (and was correspondingly rejected), and the paper benefitted (and its later publication certainly also resulted) from all the comments I got.

Anyway, how he/she phrased his/her comments (he/she must have been a big shot) while showing complete lack (of want) of understanding in detail still gets me groping for a club... You don't want a review starting as follows (as it did), do you?¹⁶

Unfortunately, the paper cannot be published. The development of ideas is premature, and the formalism and discussion is full of errors. Also the structure and presentation are lacking, thus the goals the author wants to achieve are not clear.

I should definitely publish this on <https://twitter.com/YourPaperSucks>. And by the way, I was happy and relieved to find there the letter of Hunter S. Thompson to Anthony Burgess concerning the latter's submitted text. It is as harsh in judgment, but even worse in wording. Google it, I won't cite it here.

Interestingly, I was once a third reviewer myself. Apparently, there had been a tie, and the journal needed a third opinion. I made very clear that for me, rejection was the only option. Later, however, I was informed about the decision: accepted with "Major revisions needed". And I saw that they had asked the opinion of a

¹⁶This is the case where I complained and asked the editors to reconsider, by the way.

fourth reviewer, who was more generous than me. Hm. Who wants what/whom here most?

As to the quality of the two, three, four or even more reviews let me cite Ernst Pöppel, a renowned neuroscientist, from his web page:

One of the best paper[sic] I ever published together with some colleagues had 11 reviews before it was accepted for publication.

11 reviews!

15 How it should always have ended

So far, I have elaborated on *negative* aspects of reviewing for the most part. This might give the wrong impression of my experiences, and of reviewing in general. Let me therefore make a break here and give you an example for a '(very) good' reviewing case. It's about the harshly criticized gradation paper I just mentioned.

Starting with the editor, Klaus von Heusinger, we had a nice cooperative e-mail (!) interaction during the reviewing process. His choice of reviewers was excellent, not just because of their positive ratings but also because of their competence and their constructive critique. They were the first to acknowledge the positive, new, and interesting aspects of my paper:

I think the most fascinating data discussed here are those in (13), which as the author notes have not been give much attention before

I feel that the major (and exciting) contribution of the paper is an entirely new perspective on the data

These are comments which you do not find in bad, let alone ugly, reviews.

Both reviews were thoughtful and the comments were well-considered throughout. This does not mean that I agreed with each point made, but it was a healthy mixture of assent and non-assent at any rate.

As I had to respect the reviewers' elaborate comments and to show how I had taken them into consideration, I sent a re-commented file back. By color-coding agreement/change, non-agreement/non-change and neutral discussion, I could easily show how the discussion of the comments led to an improved paper. Note that red coding for disagreement is rare here, but typically dominates in commenting bad reviews (as in the *Mind and Machines* case). I furthermore admit that my style of re-commenting adapts to the review category, and I take full responsibility for that.

16 Pathfinder: A beautiful mind

Even if it may only be true for the non-technical disciplines I am concerned with,¹⁷ scientific reviewing seems to be different from what I always thought it was and what I still think it has to be: an objectivity-oriented rather than a subjectivity-dependent procedure of securing scientific progress. Overall, it turns out to be evolutionary rather than fully determined by rationality. The papers that survive are not necessarily the best, only the best adapted.

The problem might lie in the locality of reviewing. Scientific progress is determined by local decisions (on the basis of certain disciplines, current paradigms, recent theories, given peer groups, thematically restricted journals, factual editors' choices and actual reviewers' knowledge and behavior). Therefore, reviewing as it currently is may optimize paper selection. But if this kind of search of the best path for scientific advancement were seen as a search problem of AI, then it would have to be categorized as one of the simple strategies (so-called 'hill-climbing') that run into the well-known problem of local maxima. That is, scientific progress in some field might simply come to a halt, after having been on the wrong track for a while.

This is not desirable, and one should find ways to avoid it. Intuitively, this would require more open-mindedness of some sense in scientific discussions (including reviews). In an introduction to papers of the OPEN MIND project, Thomas Metzinger and his co-author elaborate on the aspects of such a methodological stance they propose (my emphasis):

¹⁷in which the quality of a theory/model/approach cannot be measured so easily

This variant of open mindedness is characterized by epistemic humility, intellectual honesty, and a new culture of charity. It also has a pragmatic dimension: open mindedness of this kind is research generating and fosters an environment of sincere and constructive interdisciplinary collaboration. And it is profoundly inspired by the classical ideals of philosophy as a pursuit of genuine insight and rational inquiry, the importance of a critical and in a certain sense non-judgmental attitude, and the deep relationship between wisdom and skepticism as an epistemic practice.

[... This involves] reading others' statements according to the best, strongest possible interpretation—that is, to never attribute irrationality, falsehoods, or fallacies to another if alternative and more charitable readings exist.

17 Conspiracy theory: Proof

I have not yet met that much open-mindedness in reviewing and reviewers. Sometimes, one rather meets strange verbal behavior, as if produced by exponents of conspiracy theories or by UFO believers. The common characteristic of those is that they *believe* in something and try to *prove* their beliefs/theories.

There are reviewers who do just that: they believe that a paper is insufficient/bad and come up with justifications for this belief. They might, for example, tell a long story about how some aspect is treated (differently) in the literature, what they think the paper can or cannot explain, or how they don't like or understand it.

Unfortunately, such behavior is unscientific. It is one of the current axioms of science that theories cannot be proven (right), but only be disproven (proven wrong). Instead of just *stating* some negative aspect, they should *show* that at some point, the argumentation of the author is wrong for this or that reason.¹⁸

In other words, if the paper's theory cannot be disproven by the reviewer, the paper should better be treated as if it was right. Such a principle, corresponding to the *presumption of innocence* in law, could be viewed as a formal implementation of open-mindedness in reviewing: *in dubio contra rejectionem*.

¹⁸Aside from that, reviewers should distinguish the main and important aspects of the paper (core of the proposal, argumentation) from the less important ones. Of course, they should put a focus on the former. In any case, it doesn't help much to only show flaws regarding minor aspects.

18 F for Fake: Collateral damage of idiocracy

There's a lot about fake news in the news today. One should think that fake news and reviews are totally different sorts of texts: while the former are (intended to be) distributed widely, the latter are not. There may be some similarities, though.

Consider ugly reviews. For the author, the assertions made by the reviewer may not be different from the assertion that information about climate change is a hoax and that one should better burn more coal to shield earth from the hot sun. Both with fake news and ugly reviews, it is hard to understand why –despite irrational content and bad style– they still find approval by some people. Both may have more or less direct impact on non-believers of the texts.

Generally, one must be careful to prevent a change of democracies into political systems led or supported by idiots. Correspondingly, one should somehow try to prevent intellectually non-satisfying reviews/reviewing. Else, there might also be some considerable collateral damage done to the systems.

19 Trust and Shame

Ultimately, all discussion of reviewing and its problems boils down to the important role of a single bidirectional aspect: *trust*. In an ideal world, reviewers and reviewees can trust each other to be competent participants in the process, which is established/guaranteed by the editor.

Needless to say, however, that this is not the case in our non-ideal world (given the actual experience). Yet while this seems to mark a general problem of reviewing, there does not seem to be a corresponding general solution in the scientific community. For example, reviewing could be regarded and performed as a scientific discourse moderated by the editor (even double-blind with the help of capable reviewing systems) in which misunderstandings could be clarified to establish trust.

Unfortunately, such improved reviewing would probably slow down the whole process (is that so?), and therefore be in conflict with the predominant 'quantity-instead-of-quality'/'publish or perish' view of science. It's a shame.

20 Blind date

There is something else very wrong with anonymous and double-blind reviewing.

Imagine you're looking for a (new) partner and decide to use a dating portal which boasts with the advantages of information-restricted ('blind') online dating. These are the rules: your counterpart is not allowed to talk in a way that reveals his/her social, economic/financial, or health status. He/she is not allowed to tell, and you're not allowed to ask.

That is, you won't know much about the person you date, and you literally decide at the face value of what you see. Now everything can happen: you may be satisfied, or you may be dissatisfied, or you may not know whether to be the one or the other (although you will have to make a decision).

Unfortunately, the information you get (view and conversation) will probably not give you the information to decide adequately. Therefore, the candidates you dismiss could be adorable, potent, healthy millionaires, and the candidates you prefer could, although nice and good looking, turn out to be poor bipolar schizophrenics.

So you wouldn't date like that, do you? But then, assuming that double-blind reviewing ultimately works just like that: Why hasn't anyone stopped reviewing (and therefore science)'s being based on such an idiotic blind dating scheme?

21 Cinderella, Judge Dredd, Inception

Apart from all the problems we have seen in reviewing: do we know what counts as a "good paper" at all? This should be a prerequisite of what many seem to regard as the essential task of the reviewers and editors: to pick the good lentils out of the ashes.

But wait, this is a horrible misunderstanding. It would mean that in each case, a handful of people, at most, is assumed to know (and empowered to decide on) the global quality of some article, determining its further path, its impact on the scientific discourse, and the development/career of its author. Hey, we are not at the law court, and even there the matter is processed openly, arguments being presented and discussed meticulously before judgment. In reviewing, the whole procedure is somehow hidden, much more superficial, and the defense part is completely missing (!), somehow a Judge Dredd comic-book version of law (and judgment).

Actually, the task of the reviewers and editors is (or should be) much more restricted: to filter out the bad papers, not to judge whether some paper is good (in the 'objective' sense). To identify and name good and bad aspects, weigh them up and give an estimation of overall quality.

Within some range of non-badness, it should be the scientific community that determines the effective quality of some paper, and authors should be given some

leeway to plant new ideas into the reader's minds. Only then, innovative papers have a chance of being read, although they might not fit the expectations of the reviewers, or might not be fully understood by them. In the worst case, rather than being the gentle doves, reviewers otherwise might act like the nasty stepsisters preventing Aschenputtel from marrying the prince.

Unfortunately, this focus on judging is the prevailing aspect of reviewing in reality. Lately, I received the reviews of an article with the decision to rewrite the final part and resubmit. The reviewers had nodded to the first 2/3 of the article, but simply did not understand the final part in which I present my new approach to what the paper is about. Here, before going into details, I sketch the basics of my theory and point to my papers for further information. Unfortunately, they are anonymized as requested! So the reviewers were left in the dark with a new and complex interdisciplinary approach and felt insecure. This is understandable. But why should I have suddenly become a bad incompetent author in the final part?¹⁹ And why should their non-understanding be a criterion for non-publication? This would, to say the least, be a much too simplified solution for a complex constraint satisfaction problem given by the abilities of author and reviewer, the complexity of the theory and the available textual space.²⁰ The case is still open, by the way.

22 Anonymous: Perspectives

I do not think that there is an easy answer to the question of anonymity in reviewing. There were certainly good reasons to introduce (double-)blind reviewing: it was supposed to foster and secure objectivity in the process and to reduce the usual socio-emotional consequences observable in every type of critical writing. As I have tried to show, however, objectivity is neither reached nor secured, and the socio-emotional consequences have been imposed quite unilaterally on the author.

Objectively, then, there seems to be a tie in the pros and cons of anonymous reviewing. Accordingly, taking up a stance here is more a matter of practical considerations and of perspective. Disregarding the practical aspects (although they will probably tip the scales in this discussion), the main different perspectives on reviewing are either *expecting the worst*, or *hoping for the best*. Obviously, anonymous reviewing is an instance of the former.

Personally, I would generally opt for non-blind reviewing, i.e., the latter perspective. This is not easy, and there is the danger of non-perfect comments on both

¹⁹I then tried to make the part more readable and understandable, but this is far from "rewriting".

²⁰Imagine the possibilities and benefits of non-anonymous reviewing here.

sides. As to tolerating them, one of my favorite essayists proposed the following rule these days²¹:

Insults and threats are always to be rejected, criticism and mockery must be tolerated by everyone.

Yet, as I already said, there is an intermediate solution which I would ultimately prefer (considering practical aspects, and with a realistic view on open reviewing). The current situation could be much improved if reviewers and authors would be allowed and even requested to communicate with each other to clarify misunderstandings and to answer open questions. In our modern times, this could and should be implemented in an easy-to-use way (ideally, allowing Whatsapp-like textual interaction), like for example:

Rev1: p. 31: are you sure wrt. the maximal axis and the example?

Author: Yes, there's a whole theory behind it, spelled out in Anonymous (1991)

I KNOW this would help a lot.

Afterword: End of watch

Those who don't know me may ask: who is this guy lecturing us about reviews and reviewing? Let me tell you who I am.

In my fourth semester (in 1985) of my computational linguistics studies I decided to build a natural language generation system (which nobody asked or urged me to do). If anyone had known, he would have thought I was crazy. Building such a system was deemed a many-man-year venture at that time, and even then some failed. The route description generation system I successfully developed was later integrated in the big LILOG²² project and ultimately ran in a natural language system prototype presented at CEBIT, Hannover.

I then implemented the theory of a linguist who later became the director of one of the biggest linguistic institutes in Germany (the ZAS Berlin), co-wrote a paper, presented it on the German conference of Artificial Intelligence, and co-wrote a book on that topic, all that still being a student.

²¹Harald Martenstein (2017), "Über Duldsamkeit, Kritik und Spott [About forbearance, criticism, and mockery]", *Zeit Magazin* 24, p.10. My translation.

²²in a nutshell: Linguistics and Logic; full meaning: *Linguistische und LOGische Methoden zum maschinellen Verstehen des Deutschen* (linguistic and logical methods for the machine understanding of German)

My studies and especially my Master's thesis already had a general, interdisciplinary Cognitive Science perspective on natural language (processing). Working in a Computational Linguistics and Artificial Intelligence department, my first dissertation project then was about concept theories and their relevance for knowledge representation and learning. I had asked myself, if concepts are abstractions (the prevailing model at that time) and abstraction means losing allegedly unimportant (e.g., modality-specific) details, what determines pruning the details and how can there be such (context/situation) specific concepts as have been observed? I gave a presentation of this research question (plus an overview of loads of literature I had found). Nobody understood what I wanted, or wanted what I had understood (and maybe I was too provocative again), I therefore let go of that idea. Meanwhile, however, research in aspects of situated/grounded cognition (an extreme version of that point of view) has become famous, and corresponding (e.g., instance-based) models have been developed in psychology and AI (the applied versions of which are collectively known as *deep learning* today). Recently I came across a thesis from 2009 whose content came relatively close to (but in some respects still lagged behind) what I had in mind 17 years earlier.

My next (and ultimate) dissertation project was about the role of attention in spatial representation and semantics. While this was a turn back to spatial linguistics stuff I had worked on before, I had found interesting literature when scanning recent journal articles for my previous project. Unfortunately, riding such a recent scientific wave, I was probably the only one in Germany, and perhaps one of only few in the world, who saw a connection here. Accordingly, the reception of my work was disappointing. The situation has changed since, but that comes too late for me, and the majority of the cognitive scientists still has not yet acknowledged the importance of the role selective attention plays in the cognitive system. As to German Cognitive Science, what can you expect of 'experts' who in a bulletin of their society admitted that they lagged behind the American state of research as much as 15 years (at least when I cited that in my above-mentioned book review)?

When in the mid-1990ies there was a structural change to come at the university of Osnabrück, where I worked, and everyone complained about the restructuring and financial cutting to be expected, I was the only one who stood up in a meeting and pointed to the positive perspectives of that situation. When sometime afterwards my boss gave me a brochure of the DAAD about funding new international courses of studies in Germany ("take a look at that"), it was about five minutes later that I ran back to him saying that we definitely should apply. Now, Osnabrück has *Cognitive Science* as an international interdisciplinary course of studies, and a corresponding flagship institute with a 50+ people staff. What do you know!

Perhaps you can see the pattern here: scientifically, I am always a very early bird (this may be different in real life), so the worms haven't come out yet. This lack of timing leaves me starving (also because I am much too stubborn to adapt). Seriously, after years of experience in some domain, I do apparently develop some sense of research direction. I wonder why this should be different with the paper that inspired me to write this piece (see the following case report).

I'd like to end this with a striking incident. Recently, I submitted a paper that refers to Bob Dylan's work (more exactly, to *All along the watchtower*) and whose main title is *From motion perception to Bob Dylan* (it's about the semantics of directionals). Something like that is rare in science, isn't it? A few weeks later it was announced that he won the Nobel prize for literature. Pure coincidence, I have to admit.

But *All along the watchtower* is a good source for closing words of this piece.²³

*"No reason to get excited," the thief, he kindly spoke
"There are many here among us who feel that life is but a joke
But you and I, we've been through that, and this is not our fate
So let us not talk falsely now, the hour is getting late"*

Cold case: A History of Violence

If you're interested, let me talk about my recent paper (at the time of starting to write this text) again, as it gives rise to all (negative) flavors of reviewing. Imagine you have an idea that is based on long-term interdisciplinary experience and some time of research, but that is in conflict with most, if not all, disciplinary paradigms. You want to elaborate on that idea, so you need scientific feedback and funding. The question is, however, where to start.

Since I am not a conference attendee anymore (and since I had no interest in throwing myself in some disciplinary lion's den), I decided to first try some opinion

²³Some say that the song is about revolution, of lower and middle class (thief, joker) against upper class (the princes in the watchtower). In our context, this would correspond to authors and reviewers plotting against the current reviewing system. I like that interpretation.

piece to get feedback²⁴, and if positive, then to apply for funding. Since Computational Linguistics is my nominal home discipline (aside from having a corresponding degree and some practical experience, I am co-editor of the German introduction to this discipline and have written an introduction to language technology), I chose *Computational Linguistics* and its *Last Words* section. As it turned out later, the choices I made were all wrong.

The idea I have is a paradigm shifter. I argue that the most well known and widely used logic, *Predicate Logic* (PL), is defective and should be revised according to modern insights. Doing so, I am clear, simple, but specific about the single basic aspect that has to be changed according to my proposal. I was fully aware of the fact that this would question the whole paradigm, but I thought that either my argumentation would be right and convincing, or that I would be proven wrong by good counterarguments of competent reviewers.

Last Words section pieces are usually short and should offer “a personal opinion or provocative perspective on some aspect of the field”. I had an important opinion to tell (“PL is defective and needs to be revised, which is relevant for the discipline. This is quite obvious, but nobody acts accordingly!”)²⁵ and did so provocatively: here’s the intro and the last paragraph of my first submission:

Quite recently, I had a discussion with Joseph Weizenbaum, Terry Winograd, Bill Woods, and Wolfgang Wahlster. We talked about past ambitions in the field of *Computational Linguistics and Natural Language Processing*, and compared them to the current state of the art. We acknowledged the practical achievements made in recent years (see Carstensen [2013b], Carstensen et al. [2010]), but we all agreed that there is virtually a stagnation in (computational) semantics, leading to an aggravating slowdown in the progress of the whole field, and to a limitation to pattern- or statistics-based technologies.

I told them that, with all due respect to Gottlob Frege, this can be traced back to his conception of what is now known as *Predicate Logic* (aka *First-Order Predicate Logic (FOPL)*). I explained to them why predicate logic in its current form should be regarded as ill-suited for the tasks at hand and pointed out that without modern semantic fitting, *Computational Linguistics* itself might some day only be remembered as the discipline of Chatterbots and sophisticated text/speech crunch-

²⁴I had asked colleagues beforehand, but the replies were less helpful than I had expected. I could already sense some disbelief, though, and some even seemed to think that I had gotten crazy.

²⁵You must know that “deep AI”, based on logic, had dramatically failed in the 1980s, but only the scientists *using* logic –and the unfit systems– had been blamed, not logic itself.

ing. They were thrilled about this provocative perspective and asked me to write my arguments down as a *Last Words* of this journal.

I may have dreamt this, but anyway, the following is a serious attempt to motivate disposing of predicate logic in its current form.

[...]

Somehow, the reverence of the scientific community for FOPL despite its apparent deficits is reminiscent of Andersen's fairy tale *The Emperor's New Clothes* where no one dares to speak out the obvious truth. Yet what *Computational Linguistics* (among others) needs is a logic that is fit for the tasks at hand and for those to come in the next decades. So get dressed, FOPL!

As I found it unsatisfactory to only criticize, I also presented aspects of and some motivation for an alternative (which I regarded as an asset of that opinion piece). Doing so, however, this piece seemed to exceed its expected length and rather looked like a "normal" paper. Correspondingly, the editor (after discussion with a co-editor) wanted to recategorize it as a "squib". Yet this was not my intention, so I explained my point of view, insisting on a *Last Words* text status.

As a matter of fact, the editor gave in and I was granted one (!) reviewer for the adapted text which I had cut down to 8 pages and which now had already lost a bit of its original vigor. The review, however, was everything but a helpful feedback on my idea. It rather put me down as a braggart, close to scientific cheating:

However, the 'Last Words' column in CL is not intended for polemics saying that what the world needs is one's research program, of which preliminary sketches have been published here and there, but none of which has as yet resulted in anything at all new.

Hm, he missed the central point of my paper saying that what the world needs is a *different* research program. He also gets driven away by the anger about the fact that someone dares to question the holy axioms of his research paradigm, and all that polemically. It is a classic example of hubris (the irony being that it suggests hubris on my part) as the reviewer simply doesn't get the point (even in the parts where he should have been able to get it) but tries to teach me. I realized that this was leading nowhere (and that it was all my fault), sent back some quick re-comments (you can find them here, together with the pathetic little review) but accepted the rejection.²⁶

²⁶Don't get me wrong: I still think that the right reaction would have been something like 'Wow, this is interesting, let's give this some people to read and see how we can do this.' It's just that I realized that this was not going to happen.

Afterwards, I reworked the paper, extended it substantially (~40 referee-formatted pages), and submitted it to *Linguistics and Philosophy* which covers the theoretical domains of my work (semantics and logic). Suffice it to say that I attached a cover letter in which the background of the submission and a description of my expertise²⁷ was detailed. Yet we already know what happened: I got one (!) review, in which I was categorized as one of <sarcasm>those well-known AI braggarts</sarcasm>.²⁸

This review is a prime example of what can and must not be done wrong in review writing. As a matter of fact, it served in some parts as a model for my satirical reviewing rules. It is perhaps even more arrogant, superficial and inattentive to my argumentation than the one before, and full of “wicked imaginations”. I therefore did not write down re-comments carefully, but annotated the pdf in the same way I treat errors in my students’ papers. I sent the annotated pdf to T.E.D., complained a bit, and never got a reply. It contains my assessment of the review which was a delight to write:

Hence, no sufficient ground for rejection. Instead, the review has a strange start, a superfluous middle and a not so consequential end. There is too little understanding and there are too many (wrong) assumptions. Besides that, there is no word of appreciation for the ideas presented which could be the start of a discussion. Instead, the reviewer simply felt stepped on his toes and reacted unprofessionally.

Sorry, review rejected.

To complete this history of violence to my idea, I then presented the full discussion of theoretical and practical aspects in my submission to *Mind and Machines*. Again, the outcome is already known: while the two monodisciplinary reviewers at least stayed polite and acted professionally, none of them got the whole interdisciplinary picture. Accordingly, almost all of my re-comments are red. But note that one reviewer at least finds some positive words:

The author’s novel idea concerning the treatment of quantification seems interesting and worth pursuing.

"I’m the luckiest son-of-a-bitch alive"...

Despite the negative aspects, all this interaction has of course helped improving the paper to the –meanwhile serious– current version.

²⁷For example, that I have taught courses/introductions to each of the following fields/ topics/ disciplines: Artificial Intelligence (practice level), Computational Linguistics, Cognitive Science, Knowledge representation, Language technology, Linguistics (especially semantics), and Logic.

²⁸On the basis of the two reviews, statistical methods would already have projected a clear tendency here... But see Fn. 27.