School of Computing Science,
University of Newcastle upon Tyne

Safe Systems: Construction, Destruction, and Deconstruction

J. M. Armstrong and S. P. Paynter

Technical Report Series

CS-TR-830

March 2004

Published by the University of Newcastle upon Tyne,
School of Computing Science, Claremont Tower, Claremont Road,
Newcastle upon Tyne, NE1 7RU, UK.

Safe Systems: Construction, Destruction, and Deconstruction

JM Armstrong
Senior Research Associate,
Centre for Software Reliability, University of Newcastle Upon Tyne,
Newcastle Upon Tyne, United Kingdom

SE Paynter
Senior Principal Engineer,
MBDA UK Ltd,
Filton, Bristol, United Kingdom

1 Introduction

Deconstructive Evaluation of Risk In Dependability Arguments and Safety Cases (DERIDASC) is a study focussed on the language used by safety engineers in their intellectual discourse. The DERIDASC project is inter-disciplinary in the sense that it experiments with techniques from philosophy, literary theory (Eagleton 1996, Culler 1997) and semiotics (Barthes 1994, Cobley 2001, Culler 2001) to diagnose problems of language, definition, and interpretation in safety engineering. The project aims to make safety engineers re-think and improve some of their habitual definitions. The project adopts methods of textual analysis usually found only in studies of the arts and literature, although the kinds of textual studies we propose have also been influential in the discipline of law (Ward 1998, Chapter 7).

In particular, the project is applying the ideas of "deconstruction" to safety texts. Deconstruction is a term coined by philosopher Jacques Derrida (see Abrams 1999 for a short summary) and denotes the analysis of a text to reveal hidden meanings, especially those which contradict the surface message. The critical reader reads a text "against the grain", concentrating less on what the author is trying to say than on issues such as what the text tries to avoid saying (e.g. the playing down of facts that might undermine what is argued) and on what is asserted rhetorically without evidence. The idea of deconstruction is to challenge the unconscious presuppositions inherited from conceptual frameworks for thinking. Deconstruction, it can be argued, facilitates the evaluation of the text and its language as 'technologies' for thinking (Clark 2000). We hypothesize that Derrida's deconstruction, with its emphasis on revealing suppressed contradictions and paradoxes in texts, may help illuminate some of the difficulties of safety argumentation and might be a basis for new analysis techniques.

In Section 2 of the paper we offer a brief overview of the problems of textual analysis as they are represented in Derrida's work and in deconstructive literary criticism in general. Our emphasis is on how the necessarily limited scope of a text renders it incomplete entailing risks of circular justification and self-contradiction. Section 3 explains 'deconstruction' as a procedure (although a 'deconstruction' should avoid reduction to a systematic 'procedure') for the analysis of binary distinctions. It presents an example deconstruction of the distinction between *likelihood* and *severity* that is commonly assumed in standard definitions of risk (see van der Meulen 2000, p. 245 for a list of examples). In Section 4, we discuss why some of the issues raised by deconstruction are necessarily of concern in safety engineering, and indeed are already having their effects.

2 The Problem of the Text

A text is a sequential structure requiring a starting point and a conclusion. It is essentially narrative in form. Thus we require a starting point for our explanation of deconstruction. We start with a discussion of some familiar problems of authorship; but by the very act, we have already run the risk of misrepresentation. This problem of choosing a starting point for a text, of laying out the basic assumptions of an argument, or of conceiving the axioms and definitions of a formal system, are key concerns of Derrida's deconstruction.

Most authors have experienced uncertainty about how to begin a text. A well-known cliché that purports to explain the usual structure (introduction-body-conclusion) is that it should "tell the reader what you're going to tell them, tell them, then tell them what you've told them". It seems to indicate that one should begin with the desired conclusion; but if this advice were to be followed rigorously a text would contain three attempts at assertion and no argument (even then, according to deconstructive thought, each attempt to repeat 'the same' meaning would be doomed to failure).

To be accepted as an argument as opposed to a series of rhetorical assertions a text needs starting points that appear in some way more basic than its conclusions. The choice of these starting points is absolutely necessary (no text could recount everything) but it imposes bounds on what can be said in the text. Thus each genre of text (a technical manual, an essay, a novel, a poem) depends upon a reader's acceptance of conventional assumptions about 'relevance' for that genre, and additionally, any particular restrictions requested by the author of the text at hand. There is a tacit agreement that certain subjects are to be considered 'irrelevant' to that genre of discourse, and will not be raised in it. This is an entirely necessary procedure, but unfortunately it is hard to argue that 'relevance' defines anything other than a socially-constructed boundary which the reality of events need not respect.

The ‘starting point’ (Derrida’s term is ‘origin’) for any text or argument cannot be absolutely justified, since one knows that reality exceeds the bound it constitutes: before the argument begins, everything else has already ‘begun’ (Bennington & Derrida 1993). So when we consider what could determine a choice of origin(s) (of basic assumptions, or of founding axioms and definitions) it becomes apparent that the most attractive criterion is that they allow the desired conclusion to be reached by some means. Thus we encounter the problem of ‘constructivism’. Origins of all kinds can be challenged on the basis that they are constructed in order to guarantee and preserve a desired conclusion, giving it the appearance of something that has been ‘justified’; indeed, the basic assumptions may even be ‘reverse engineered’ from the desired conclusion.

Although it is possible for an author to start with some unusual or important observation and work through its possible consequences with no clear end in view, this is unlikely to lead to a highly structured and disciplined text (although literary authors have experimented with the approach). Nonetheless, most texts are constructed so as to mimic this process and in a successful text the conclusions will appear to follow naturally and obviously from the assumptions. However, this structure rarely corresponds to the process by which the text was really constructed especially when the writing of the text *demands* at least some idea of the desired conclusion. Even a text that sets out to answer a question in as unbiased a fashion as possible will tend to presuppose a fixed set of expected (or already extant) answers. Indeed, to ask a question in a text is already to have made a proposition regarding the significance of the question and the possibility of an answer.

One might object that in rational argument, which proceeds according to the laws of logic and of grammar, the process by which the text has been constructed does not matter. So long as the conclusions follow from the assumptions according to the laws and so long as the assumptions are accepted, the conclusion must be accepted. However, this is also true of a tautology, which raises the issue of how to distinguish tautological argumentation from ‘meaningful’ argumentation. For example, logicians have long realised that formal argumentation and proof cannot eliminate the possibility of circular reasoning. This is partly because no logic can prove that its own axioms and definitions are indubitable truths; but it is also because circular reasoning is not a problem of logic but a problem of belief (Cambridge 1999, p. 144).

We often accept a set of assumptions in a provisional sense ‘for the purposes of argument’, in order to initiate communication, or just to get through the task of reading a text; but why should we ever be more committal and accept assumptions as true thereby accepting the conclusion? Ironically, one possible reason for accepting assumptions is that the conclusion that follows from them is already believed, or at least strongly desired. Thus even the most rigorous argument cannot eliminate the risk that its assumptions have become ‘detached from’ the reality they are supposed to describe. Biases, perceptual limitations, and the desires of the arguer are important factors; but the necessity of structuring reality into a tractable mental model is the real culprit. Indeed, deconstructive thinkers argue that the very nature of the 'signs' we use to represent reality and their irreducible role in perception and thinking ensure that abstract arguments can never be unproblematically 'attached to' reality.

Deconstruction sees the relation between language and reality as a problem of 'effects of meaning' rather than one of logic. Questioning the meaning of assumptions is of course essential to any assessment of their validity. However, in *Limited Inc* Derrida shows that any attempt at an objective assessment of 'meaning' is bound to encounter highly inconvenient obstacles (Derrida 1977). When we consider 'effects of meaning' we find that they are inherently unpredictable. They are partly determined by socially accepted rules for communication, and partly determined by the individual intentions of the communicators; but they also depend upon what is rather vaguely referred to as 'context'. To complete the definition of a meaning effect, we would need to capture 'context' in a description; but context appears to have no definite bound; and furthermore, its description merely produces more 'text' which requires context to explain it in turn. One can try to arrest the indeterminacy of context by appeals to "what a speaker (author) must have meant at the time"; but this is another disguised appeal to context. It is complicated, for example, by the question of how unconscious mind relates to conscious intentions, and whether speakers (authors) really had clear intentions *in the context* of utterance (writing). Because of the apparent necessity and simultaneous impossibility of an objective definition of a context, nothing is more than provisionally sufficient to bound the effects of meaning that a particular text can have; these are, after all, a function of the future contexts in which the text will be repeated (read). There is therefore no 'final interpretation' of any text, even the most simple of sentences. At best we can expect meanings to be socially constructed and agreed; but this is hardly an objective foundation for rational thought.

It is common to represent the problems of natural language meaning in terms of 'ambiguity'. Derrida (1981) argues that this actually underestimates the extent of the difficulty, since it presupposes that a text has a fixed set of meanings that can be identified in principle, even if no objective choice can be made between them; but reading a text is always 'productive'. Because context itself is unpredictable and because we are condemned to shift through time, any text retains the potential for effects of meaning that cannot be foreseen. Each time we re-read a text we may notice (produce?) something new. As Derrida puts it, "iterability alters" (Derrida 1977, p.62); and it does so because previous contexts can never be perfectly described and are never exactly reproduced.

Wheeler (2000) summarises Derrida's thought as a working out of the consequences of there being no 'magic language'. A 'magic language' would be one to which all other languages could refer for their 'literal meaning'. Usually, we think of 'thought' or 'perception' as the things to which our utterances refer; but Derrida argues that these too involve the manipulation of signs, and are therefore linguistic, with all the interpretive insecurities and indeterminacies that natural language exhibits.

The problematic nature of the idea of a 'magic language' and of the opposition between 'literal' and 'figurative' meaning may not be immediately apparent; but it suffices to consider whether the 'literal meaning' of a statement could ever be written down non-figuratively in some language. If such a language existed we would be well advised to discard all others. Its meaning would never be in doubt, and it would provide an absolutely secure basis for thought and expression. All of poetry with its complex metaphors would be expressible 'literally' in it. The very fact that we can translate texts from one language to another seems to suggest that there is indeed some 'literal' meaning being transferred; but Derrida counters that the process of translation is always somewhat imperfect with respect to the source text and creates new meaning effects in the target text.

Since the late 1960's, Derrida has published around twenty major works arguing that philosophy has persistently resisted close consideration of the unpredictabilities of writing, of representation, and of interpretation because of the disconcerting implications for thought, rationality, and logic. If meaning effects are essentially unpredictable, we cannot be sure that we are thinking 'rationally' at all. He argues that the idea that meaning can be mastered, described, and bounded is an unreasoning prejudice which Western modes of thought secretly depend upon. He terms this prejudice 'logocentricity'. For Derrida, since a text is 'founded' insecurely on its meaning effects it is a rather perilous structure, condemned to conceal knowledge of its internal contradictions whilst also being condemned to reveal them to readers who will not play the author's game. Since it is in the nature of any 'sign' to be arbitrary with respect to its referent, there is only a conventional, not a natural, attachment between them. So when considered as a network of cross-referring signs, a text always has the potential to mean something other than what the author intended and perhaps even to contain a *denial* of what the author was trying to say.

Derrida's deconstructive analysis of the dilemmas of the text can be loosely compared to Godel's discoveries about incompleteness in formal logic. A text can attain an appearance of consistency only by self-imposed bounds on its completeness; as the text attempts a more complete treatment of its subject, internal inconsistency, loss of structure, and lack of conclusiveness (Derrida uses the term 'closure') become more apparent. A deconstructive reading concentrates on the way that terminology and criteria of relevance are 'constructed' in order to try and prevent a loss of control over meaning, whilst maintaining the appearance of an ordered progression from assumptions to conclusions. Thus instead of refuting the arguments in a text, a deconstruction concentrates on revealing the limits of authorial control. The challenge is that an apparently coherent, logical, and structured text achieves its ordered appearance not by representing truths, but by suppressing anything that contradicts or undermines its intended message or else renders that message trivial in its own terms. Deconstructive readings often seize upon issues mentioned only in passing in a text and attempt to show that when the implications of these issues are followed through rigorously they contradict the main line of argument.

The alleged inevitability of this kind of internal collapse derives from the fact that a text needs starting points - ‘origins’. Deconstructive thinkers argue that since a starting point is a necessary imperfection, it is inevitable that in the course of thinking and writing ideas that cannot be properly accounted for in terms of the argument will crop up. How can a text deal with such ideas? It might resort to a kind of ‘myth of The Fall’ in which the valued origin is somehow tainted from without; it may fall silent at what appears to be a crucial point; it may try to evade the difficulty by subtle appeals to the readers background knowledge, culture, or charity; or it may resort to blatant rhetoric in order to divert attention from inconvenient facts, contrary views, or deep uncertainties; misrepresentation or caricature might be used in order to weaken the force of a contrary idea or opinion. Rhetoric is commonly used to assert that certain issues are irrelevant to an argument, which raises the question of how they came to be raised there in the first place.

All these phenomena are indicative of problems with the orderly progression of argument in the text. Indeed, the argument may not be ‘progressing’ at all; perhaps the desired conclusion is merely trying to re-assert itself in different terms, and since “iterability alters’, inevitably failing to do so.

More positively, Derrida is interested in how, at times when argumentation and terminology falter, valuable progress can be made by a resort to analogy and metaphor. In rational discourse the status of figurative language is problematic; metaphor is regarded as an unsatisfactory detour on the way to truth; but deconstructive thinkers are generally positive about figurative language, and adopt Nietschze’s dictum that ‘truths’ are merely metaphors whose status as such we have forgotten out of habit (for a discussion see Gayatri Spivak’s introduction to Derrida’s *Of Grammatology* – Derrida 1967). Another a-logical but quite productive strategy is the introduction of a neologism. Neologisms usually have recognisable roots in familiar words; indeed an author may reuse a familiar word without alteration, thus extending its meaning into new contexts. Derrida calls the study of this process ‘paleonymics’. These processes produce new meaning effects whilst drawing on the power of existing words. Such strategies are not risk free. Neither are they obviously justifiable by any criteria except practical necessity. Yet without them argumentation would be exceedingly difficult, and perhaps even sterile or impossible.

3 Deconstructing Risk

The usual starting point for a deconstructive reading is the analysis of the binary distinctions that a text depends upon. Deconstructive thinkers claim that each term of a binary opposition will ‘contaminate’ the other. This view arises from a theory of language according to which the meaning of a term is determined by its position within the linguistic system, and not by any fixed property of ‘meaning’ that is indissociably bound to it. A ‘meaning’ is an effect produced by the inter-relationships among the terms of a language. Consequently, neither concept in an opposition of contrast has an identity that is entirely independent of its ‘opposite’. For example, if we take one of the terms of an opposition and try to define it, we find that we can only do so by mentioning the other term, and vice versa. Each term contains what Derrida calls the “trace” of its opposite: so in deconstructive thinking, concepts are impure, or to put it another way, distinctions can always be undone by abstracting something that is common to both of them. Derrida claims that the usual result is an explosion of complexity of discriminations, rather than mere vagueness (Derrida 1977).

As an example we will examine the distinction between the *likelihood* (probability) of an accident and the *severity* of an accident. These two terms are broadly accepted as independent variables which combine to make up what we call ‘risk’(van der Meulen 2000, p. 245). It might not be thought that either of these terms could be seen as more ‘original’ or ‘valuable’ than the other. Yet experts in probability theory tend to argue that however severe the consequences associated with a risk, if the probability can be made so small that the overall product (the total ‘risk’) is insignificant, then risk acceptance will be rational. Those who distrust probability theory argue that probability models are prone to give us the answers that we would prefer to hear, so that the severity of the possible consequences should be the primary criteria in risk acceptance. We can see this debate in terms of a disagreement about which component of risk should be of prime importance, even though the definition of risk as their ‘product’ seems to suggest that neither is primary.

The next deconstructive question will be to consider how severity and probability could be interdependent. They are presented as independent variables, but a close questioning of this assumption quickly complicates it. A measured probability of an accident, against which any estimate of likelihood is ultimately assessed, cannot be determined until the system lifetime has passed. However, this ‘objective’ measure can itself be determined by the severity of accidents. For example, suppose that an airliner crashes on its first passenger flight. Consider two possible futures: a) political pressure keeps the aircraft flying and it goes on to build up an enormous number of accident-free flight hours; b) the crash is used as a reason for cancelling the aircraft programme. The different lifetimes make the “objective” probability of an accident per flight hour very different in each case; but the aircraft lifetime is dependent upon the political will to either keep flying the aircraft or withdraw it. Concorde provides an example of this. A recent crash transformed Concorde from one of the safest aeroplanes in terms of accidents-per-flying-hour into one of the most dangerous (Daily Telegraph 2000); but the political will existed to keep Concorde flying. If it had not, Concorde would have remained a statistically ‘dangerous’ aircraft. It is possible that Concorde will go on in future to build up further flight hours without incident, thus apparently ‘becoming safer’. The idea that a measured probability is ‘objective’ is undermined when we realise that the size of the sample space from which the measurements were taken is determined by something that is not an objective given. As far as statistical measures of accidents-per-unit-time are concerned, the size of the sample space is determined by willingness to carry on living with the severity of any actual consequences. Thus, not only are probability *estimates* dependent upon severity estimates - in the sense that we may unconsciously underestimate the probability of events we fear – but even the *measured* probability of an accident-per-unit-time is dependent upon its severity should at least one accident actually occur.

Further analysis also produces the explosion of complexity of meaning predicted by Derrida. We have argued that what one might think of as "objective" measures of risk *likelihood* depend upon subjective willingness to live with risk *severity.* However, we must further ask what determines our willingness (or unwillingness) to live with the possibility of severe consequences and why we might continue an activity even after it has already had severe consequences. Evidently, this question leads us to a consideration of enormously complex (even imponderable) issues of choice. For example, often in the aftermath of an accident we are led to reiterate the original question of risk acceptance. This may happen even where the accident falls within the expected probability of occurrence. If, after a (re)consideration of the likelihood of recurrence and the severity of the consequences, we make changes to the system to reduce future risk, we cannot avoid the challenge that we have retrospectively invalidated the original risk acceptance argument and that any new argument will be just as fragile as the first one. The rather-too-neat distinction between *likelihood* and *severity* perhaps (mis)represents risk as a simpler, more empirically verifiable, and more manageable concept than it really is. Yet it seems impossible to do without this distinction. Indeed, the 'deconstructive argument' given above itself depends upon it. Culler (1998, p.149) captures this paradox in an amusing way, noting that deconstruction can be described as "sawing off the branch upon which one is sitting".

4 Relevance To Safety Engineering

Derrida denies that deconstruction can be reduced to a corrective procedure or method, preferring to define the unravelling of conventional meanings merely as "what happens". Therefore, we do not propose the deconstructive analysis of texts as a replacement for the more usual procedure of analysing safety arguments for flaws in the reasoning, inaccurate data, and gaps in the evidence; but neither do we consider them as necessarily 'secondary' in relation to such tasks. Were we to do so, we would fall prey to a form of contradictory logic that Derrida (1967) calls 'the logic of the supplement'; for example, if we state that the textual analyses we propose are mere 'supplements' to current procedures, then we imply that they are both necessary *and* unnecessary to current procedures. Anything that can be 'supplemented' must by definition have a basic deficiency or lack that the supplement can remedy; there is therefore no reason to consider it as 'primary' or self-sufficient in the absence of its supplement. Deconstruction must already be going on in safety engineering, but not under that name. In this section we argue that this is indeed so.

The problems of pinning-down intended meaning, of finding implicit assumptions and circularities, and identifying what is a-logical, or even incoherent in a text are of interest to anyone involved in assessing safety texts. The text, with all its risks, is a technology that safety engineering seems unable to do without. Indeed, we are not the first safety experts to take an interest in deconstruction. For example, Turner (1994) discusses Derrida's ideas about the relation between chance and necessity, and how it relates to the interpretation of accidents. On DERIDASC we have turned to the textual representation of rational argument and rational choice in the process of safety acceptance.

For texts that present safety arguments the conclusion in view is specified in advance and necessarily so: naturally, the purpose of a safety argument is to arrive at the conclusion that a system is adequately safe. Once the argument has been constructed (usually by the supplier of a system) so as to reach this pre-specified conclusion, it is analysed for flaws according to whatever regulatory or assessment criteria happen to be in place in the application domain. The rules of this exchange seem simple enough then: an argument is constructed; it is put forward as valid; then it is tested to 'destruction' in the sense that flaws of logic, gaps in evidence, and inaccurate assertions are identified. If no flaws are found, the argument has survived its tests and the system in question will be provisionally accepted for use.

However, most safety experts will recognise this description as extremely idealised. It does not recognise phenomena such as: ambiguities in the interpretation of evidence; professional disagreements about what constitutes 'best practice'; collectively recognised limitations (of knowledge, of technology, of the intractability of certain problems). Neither does it recognise the political, legal, or moral issues involved in safety acceptance; nor changes in public opinion about risk acceptability; nor indeed the question of how (or whether) public opinion is to be elicited in the first place. When so many different forces can be brought to bear on a particular risk decision it seems unlikely that an argument solely based on rational calculation will be sufficient to determine the decision, especially if, as deconstruction implies, the 'meaning effects' of the calculation (as opposed to its more tractable mathematical meaning) will be at the very least problematic.

The challenge that apparently 'objective' models of risk decision-making are really post-hoc rationalisations of asserted biases is not unprecedented in safety engineering. For example, Adams (1995, Chapter 6) argues that since there is no objectively meaningful notion of 'value' upon which to found the monetisation of risk, the results of Cost-Benefit Analysis are determined by the biases of whoever is conducting the analysis. In his words: "cost-benefit analysis is almost always used not to make decisions, but to justify decisions that have already been made".

Similar arguments have been made concerning the use of reliability theory in safety assessment. On this subject, Leveson (1995, p. 168) quotes DT Lowe: "Risk analysis of the type considered here is to safety what the merry-go-round is to transport. We can spend a lot of time and money on it, only to go round and round in circles without really getting anywhere."

Furthermore, problems of constructivism in interpretation often emerge in the context of discussions about Probabilistic Risk Assessment. Crawford (2001) reviews objections to PRA ranging from the subtle to the unsubtle (quoting physicist Richard Feynman as saying "If a guy tells me the probability of failure is 1 in 10^5, I know he's full of crap."). A rejoinder to Crawford's paper provides an amusing example of how different representations of a problem can lead to mutual incomprehension, circular self-justification, and unconscious self-contradiction. Vesely and Fragola (2002) object to Crawford's questioning of the meaningfulness of statistical testing with the words: "The author's conclusions may be interesting conjectures, but they have no statistical basis." This misses Crawford's point entirely, especially as Vesely and Fragola agree that: "*No* PRA should be taken at face value" (our italics).

The phenomenon of circular justification also troubles discussions about risk 'tolerability' (Health & Safety Commission 1998). Our oft-discussed principle is that risks must be reduced to a level that is "as low as reasonably practicable" (ALARP). However, ALARP is based on the assumption that for a particular type of safety system a "tolerable region" actually exists; this guarantees that at least some level of effort will enable it to be attained, but this presupposition is never questioned. The notion of 'tolerability' implies that what is tolerated is undesirable, but in some sense unavoidable; but if risk taking were entirely unavoidable it would not require 'justification'.

We can observe analogous difficulties in discussions about the use of Commercial Off-The-Shelf software in critical applications, particularly where the software in question was not originally developed for critical use. A text that presupposes the possibility of the safety-critical use of COTS and uses this presupposition as the starting point for its arguments will find it impossible to account for phenomena that indicate why COTS use was previously discounted in safety critical systems. For example, one text we have analysed proposes a method for justifying the use of "Software of Unknown Pedigree" (SOUP) in safety-related systems. It alludes briefly to the possibility that 'Easter Eggs' might be buried in COTS software, and comments that "such problems are harder to deal with than in bespoke software …"; but the text seems unwilling to go much further. An alert reader might wonder how a software engineering manager should "deal with" Easter Egg software. Obviously, identification of the offending source code is the first pre-requisite. Subsequently, removal will be desirable. However, this could prove tricky if there are non-functional dependencies between the Easter Egg software and the useful software. It may be that (e.g. due to memory mapping sensitivities) the Easter Egg software cannot be removed; but it could perhaps be rendered harmless by a suitable "wrapper" function.

What of the development process that allows Easter Egg software to get through compilation and build into a shipped delivery? Evidently, the programmers responsible need to be identified and sanctioned. The development process needs to be re-examined and modified to make sure the childish trick will be detected if it is played again. Hence, there is no way to "deal with" a development process that does not permit the identification of individual programmers and is not amenable to the necessary preventative measures. The text avoids detailed consideration of the Easter Egg issue, but when the issue *is* followed up, it quickly leads to general principles that were well-accepted *before* the critical use of SOUP was ever proposed: e.g. that the software source code should be available for inspection; that object code and memory maps should be available just in case; that the software should be open to reconfiguration and recompilation; and that a software development process needs to be traceable and self improving. These principles indicate why, until quite recently, the reuse of SOUP in critical systems was not countenanced. A text that does countenance it can have its 'starting point' 'deconstructed' via the construction of some 'earlier' starting point.

Note that no text can evade this manoeuvre. The text in question gives various hints that the authors were fully aware of the objections to SOUP discussed above at the time of writing; for example, they have included warnings about Easter Egg software in their appendices. However, given the logic of their text and its starting point, they could not make them explicit in the main body. Deconstruction predicts that the author is always to some extent a prisoner of the requirements of their text; the arguments in a text (and this paper can be no exception) are unavoidably shaped by the starting points that have been chosen or specified in advance.

5 Conclusion

A deconstructive reading will be very alert to a text's unwillingness to follow up the implications of avowedly peripheral issues that it does not avoid mention of. There is no reason why safety arguments should not be confronted with the same challenge, particularly as the risks of incompleteness, of 'confirmation bias', and the necessity for creative 'safety imagination' when identifying hazards are already recognised by safety engineers. At the same time there is increasing recognition of the irreducible subjectivity involved in safety judgements and concern over whether 'expert opinion' is enough to cover gaps in meaningful data (Redmill 2002a), (Redmill 2002b), (Adams 1995). Indeed, the controversy that opened up during the drafting of the Royal Society Study Group report on *Risk Analysis, Perception, and Management* (Royal Society 1992) has not receded. Those who think risk management can be given a rational and scientific basis, and those who doubt that it can, seem unable to find arguments that can convince those predisposed to the other view (Hood & Jones 1999).

The deconstructive perspective would view this debate as an inescapable 'social text'. Deconstructive thought predicts that safety engineers will be unable to define any objectively meaningful unit of risk - since meaning is a relational property, there is no 'objectively meaningful' term in any case. Neither will there be any unarguable criteria for determining risk 'tolerability'. Yet we cannot abandon this impossible search without relinquishing our discourse to a self-destructive and indeed contradictory relativism. Rational safety argumentation *requires* basic definitions, but by that very fact remains perpetually vulnerable to a charge of presupposing a conceptual foundation that - if we recognise the challenge of deconstruction - it cannot possibly have. Foundational concepts - for example, the usually unquestioned likelihood-severity distinction - must be artificially constructed; we cannot base them on the 'discovery' of some truth that lies beyond all problems of interpretation.

Deconstruction can be viewed provisionally as a philosophical framework for 'reading between the lines'. It stimulates the readers' imagination by encouraging a close questioning of basic terminology and its effects upon conceptual thinking in what are, after all, rather dry technical documents. Most of the flaws sought out in deconstructive readings are not qualitatively different from those targeted in critical reading generally: e.g. ambiguities, logical fallacies, and gaps in reasoning. Thus deconstructive reading could be used by assessors to analyse safety arguments and used by systems developers in order to test and improve their safety arguments before assessment. However, our examples will have alerted the reader to the fact that deconstructive reading can undermine even the most logically 'watertight' argument by pointing out the limits of the language used to express it. A major question for us is whether such fundamental questioning leads to intellectual paralysis (in what is after all a decision-making process) or whether the insights gained will be intellectually stimulating. Arguably, the point of intellectual paralysis marks where safety 'decision-making' really begins (afresh?).

Acknowledgments: This work was supported under EPSRC research project GR/R65527/01. The authors would like to thank Bob Malcolm for introducing us to the work of Barry Turner on deconstruction and safety engineering.

6 References

Abrams MH 1999. *A Glossary Of Literary Terms*, Seventh Edition, Harcourt Brace College Publishers, ISBN 0-15-505452-X.

Adams J 1995. *Risk*. Routledge, ISBN 1-85728-068-7.

Barthes R 1994. *The Semiotic Challenge*. Translated from the French edition *L'aventure Sémiologique* (Editions de Seuil, 1985) by Richard Howard, University of California Press, ISBN 0-520-08784-4.

Bennington G & Derrida J 1993. *Jacques Derrida*, University of Chicago Press, ISBN 0-226-04262-6.

Cambridge 1999. *The Cambridge Dictionary of Philosophy*, Second Edition, Cambridge University Press, ISBN 0-521-63722-8.

Clark T 2000. *Deconstruction and Technology*. In: Royle 2000, pp. 238-257.

Cobley P 2001. *The Routledge Companion to Semiotics*. Edited by Paul Cobley, Routledge Taylor & Francis Group, ISBN 0-415-243149.

Crawford J 2001. *What's Wrong with the Numbers? A Questioning Look at Probabilistic Risk Assessment*. Journal of System Safety, Vol. 37, No. 3, Third Quarter 2001.

Culler J 1997. *Literary Theory: A Very Short Introduction*. Oxford University Press, ISBN: 0-19-285383.

Culler J 1998. *On Deconstruction: Theory and Criticism After Structuralism*. Routledge, ISBN 0-415-04555-X.

Culler J 2001. *The Pursuit of Signs*. Routledge Classics, Routledge, ISBN 0-4152-5382-9.

Daily Telegraph 2000. *So Just How Safe Is It To Travel by Plane?* Article by Matt Ridley, Wednesday, July 26th 2000. Also available at:
http://www.smh.com.au/news/specials/intl/concorde/conair21.html

Derrida J 1967. *Of Grammatology*. Translated by Gayatri Chakravorty Spivak, The John Hopkins University Press, Corrected Edition 1997, ISBN 0-8018-5830-5.

Derrida J 1977. *Limited Inc*. Edited by Gerald Graff, Northwestern University Press, Evanston IL, ISBN 0-8101-0788-0.

Derrida J 1981. *Dissemination*. Translated from the French edition *La Dissémination* (Editions de Seuil, 1972) by Barbara Johnson, Athlone Contemporary European Thinkers, ISBN 0-485-12093-3.

Eagleton T 1996. *Literary Theory: An Introduction*, Second Edition, Blackwell Publishers, ISBN 0-631-20188-2.

Health & Safety Commission1998. *The Use of Computers In Safety-critical Applications: Final Report of the Study Group on the Safety of Operational Computer Systems*. HSE Books, ISBN 0-7176-1620.

Hood C & Jones DKC 1999. *Accident and Design: Contemporary Debates in Risk Management*, Routledge, ISBN 1-85728-598-0.

Leveson NG 1995. *Safeware: System Safety and Computers*. Addison Wesley, ISBN0-201-11972-2.

Redmill F 2002a. *Risk Analysis – A Subjective Process*. In: *The Engineering Management Journal*, IEE Publications, April 2002, pp. 91 – 96.

Redmill F 2002b. *Exploring Subjectivity in Hazard Analysis*. In: *The Engineering Management Journal*, IEE Publications, June 2002, pp. 139 – 144.

Royal Society 1992. *Risk Analysis, Perception, & Management*, Report of a Royal Society Study Group, The Royal Society, London ISBN 0-85403-467-6.

Royle N 2000. *Deconstructions: A User's Guide*. Edited by Nicholas Royle, Palgrave, ISBN 0-333-71761-9.

Turner BA 1994. *Software and Contingency: The Text and Vocabulary of System Failure?* Journal of Contingencies and Crisis Management, Vol. 2 No. 1, March 1994.

van der Meulen M 2000. *Definitions for Hardware and Software Safety Engineers*. Springer, ISBN 1-85233-175-5.

Vesely W & Fragola J 2002. Untitled article in the *From Our Readers* section. Journal of System Safety, Vol. 38 No. 1, First Quarter 2002, pp 5 – 6.

Ward I 1998. *An Introduction to Critical Legal Theory*. Cavendish Publishing Limited, ISBN 1-85941-348-X.

Wheeler SC 2000. *Deconstruction As Analytic Philosophy*. Stanford University Press Cultural Memory in the Present Series, ISBN 0-8047-3753-3.