24

Imagine: Thought Experiments in Information Systems Research

L. D. Introna
Department of Information Systems
London School of Economics and Political Science
London WC2A 2AW, England
Fax: 0171 955 7385
E-mail: L.INTRONA@lse.ac.uk

E. A. Whitley
Department of Information Systems
London School of Economics and Political Science
London WC2A 2AW, England
Fax: 0171 955 7385
E-mail: E.A.Whitley@lse.ac.uk

Abstract

In this paper, we will argue that thought experiments can play a significant role in qualitative information systems research. We show the unique role that thought experiments can play in destroying existing belief systems within a community as well as how they can help creating new ones. Because thought experiments have to rely on existing data and concepts, they are particularly effective at providing the shift in perspective needed for a scientific revolution. In the paper, we analyze four thought experiments, relevant to information systems, to show how they are able to bring structure to a muddled discourse in a way that empirical, quantitative research cannot. We conclude with a discussion of the conditions necessary for effective thought experiments that will enable them to be convincing and challenging. In so doing, it is hoped that the result will be further clarity in the types of questions and answers that we should be exploring in the study of information systems.

1 INTRODUCTION

> *Thought experiments have more than once played a critically important role in the development of physical science.* Thomas S. Kuhn

Few would dispute the convincing power of a well designed thought experiment. Its ability to persuade seems to be in the compact, crisp nature of the narrative, or in the use of common everyday terms and concepts. As a result, thought experiments tend to have an appeal way beyond the limitations of disciplinary boundaries. Yet, these same characteristics also tend to trivialize them. They are often seen as nice "stories" to entertain students or add some spice to an otherwise dull lecture. While this may be true, however, there is no doubt that thought experiments have often played decisive roles in creating clarity in a confused and muddled discourse, often creating powerful counterexamples to dislodge, or seriously question, a prevailing theory. Thus, although there are many reasons to believe that thought experiments do not belong in "serious" science, there seems to be equally sufficient reason to believe that they not only belong, but could play a decisive role in advancing understanding in situations where contradiction and confusion exist (Kuhn 1977).

In this paper, we will show how thought experiments can play a significant role in information systems research and how they can be used to change the beliefs of information systems researchers. We will demonstrate, through the analysis of four thought experiments, that they have the potential to bring clarity to a muddled discourse in a way that empirical research often cannot. We will also put forward some guidelines of when and how to apply thought experiments.

Our intention is to provide a background understanding of thought experiments and the role that they can play to enable the information systems discipline to take and develop the idea of thought experiments to further the field.

The paper is structured as follows: first, we will define the notion of thought experiments and present a taxonomy for classifying them. We will also argue, using the work of Kuhn, why we believe them to be important to current information systems research; second, we will analyze four thought experiments to demonstrate their usefulness to developing our understanding of computer based information systems; finally, we will provide heuristic guidelines for the use of thought experiments in the future.

2 ON THOUGHT EXPERIMENTS

2.1 A Taxonomy of Thought Experiments

Thought experiments are experiments that are unrealizable (Bunzl 1996). They are unrealizable as a matter of principle or of practice. For example, Einstein's (1949 p. 53) observer traveling at the speed of light is unrealizable in principle. Many moral issues are similarly raised by considering situations that could not be done in practice, especially not just to make moral/philosophical points.

One may ask whether a thought experiment is not merely a particular form of argument or even a form of scenario analysis. There are obviously some similarities with these; however, the essential difference is in the fact that a thought experiment is explicitly constructed, within the confines of the existing paradigm, *in order to* destroy the existing paradigmatic position, or in order to construct an argument for a new position. One may define a thought experiment as follows:

> It is a coherent narrative of an unrealizable experimental situation, commensurate with the current paradigm, that is explicitly constructed in order to destroy the current paradigmatic position or to support an emerging paradigmatic position.

Thought experiments are not done mentally purely for the sake of convenience. They are thought experiments precisely because they cannot be performed, in principle or in practice, in the empirical world. This ought immediately to raise a number of questions. If they cannot be performed in the empirical world to what extent are their conclusions relevant to the empirical world? Is this not a form of pure speculation? Some attempts at answering this point are provided by Bunzl (1996), Sorensen (1992) and Wilkes (1988), but a detailed discussion is beyond the scope of this paper. However, in this paper we will show, in a less technical way, that it is true that certain thought experiments are relevant to our knowledge of the world. We will do this by combining a taxonomy of thought experiments with the work of Thomas Kuhn on scientific revolutions.

A first step toward understanding the range of roles that thought experiments can play will be by presenting a taxonomy of thought experiments. Brown (1986) argues that thought experiments can be either destructive or constructive, as indicated in Figure 1. Destructive thought experiments can be thought of as directed against an existing theory and are designed to undermine the theory by demonstrating either an inconsistency internal to the theory or its incompatibility with other background beliefs. In contrast, constructive thought experiments are directed at providing support for a contested existing theory or argument (conclusive), directed at clarifying a known and emerging argument (elucidative), or can be an occasion for speculation to open up the debate concerning an emerging theory or argument (conjectural).

The discussion of the taxonomy above indicates that thought experiments may have a unique role to play in changing the shape of perceptions and beliefs within scientific research. The next section will briefly review the work of Thomas Kuhn on scientific revolutions and suggest the role that thought experiments can play in them.

2.2   Thought Experiments and Scientific Revolutions

Kuhn's historical review of scientific stability and scientific change brought the term paradigm shift into popular discourse. To him, a paradigm is much more than just the methods used by a community of researchers. A paradigm is a way of looking at the world. It is a way of deciding what is an interesting question and it is also a way of deciding what is an appropriate answer.
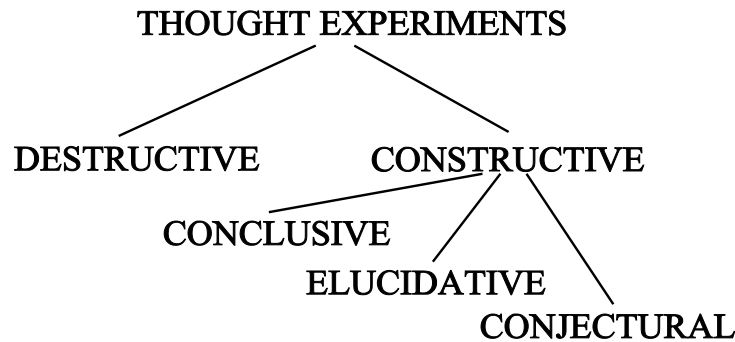
THOUGHT EXPERIMENTS

DESTRUCTIVE   CONSTRUCTIVE

CONCLUSIVE

ELUCIDATIVE

CONJECTURAL

Figure 1 A Taxonomy of Thought Experiment (adapted from Brown 1986).

According to Kuhn, scientific traditions evolve but they do so in discrete rather than continuous steps. Over time, the existing paradigm tends to generate more and more anomalies that cannot be explained within the current framework of beliefs. These problematic results are initially ignored or are attributed to errors in measurement or extraneous variables. At some point, however, they cannot be ignored and, at this point, a scientific revolution occurs and there is a shift to a new paradigm.

It is important to note, however, that a paradigm shift can only occur if there is an alternative emerging paradigm available. This paradigm must do more than just exist, however. It must be able to "explain" the majority of the existing "understood" phenomena as well as a good proportion of the anomalies. The question arises, however, as to how such shifts occur.

Paradigm shifts do not normally arise from spotting new relationships among a large set of anomalies. Instead, as the number of anomalies grows, attention becomes focused on the anomalies themselves and the beliefs that caused the anomalies will come under investigation. Thus, paradigm shifts change the very nature of the scientific work being undertaken. This is normally done through a radical revision of existing beliefs about how the world functions, as seen, for example, in the shift from an earth-centric to sun-centric view of the cosmos by Copernicus.

It thus becomes apparent that the facts in the scientific discipline are not acontextual truths but rather that they are elements of closely interrelated networks of beliefs (Wittgenstein 1969) developed within the scientific community. Thus the scientific beliefs associated with the paradigm shape what are considered to be the facts within the paradigm. In these situations the thought experiment enables the scientist "to use as an integral part of his knowledge what that knowledge had previously made inaccessible to him" (Kuhn 1977 p. 263). By shifting the paradigm, the same

problematic phenomena are now viewed with a different conceptual apparatus which emphasizes different aspects of the problem; the viewpoint on the situation has changed.

It seems plausible, therefore, to map the stages of paradigm shift onto our framework for thought experiments. Destructive thought experiments are designed to make explicit the beliefs of the existing paradigm in such a way that they unfreeze (Lewin 1952) the existing situation. Thus, destructive thought experiments play a role in starting a scientific revolution. Constructive thought experiments, thereafter, play a similar role in providing support for the newly emerging paradigm which can then lead to a refreezing of beliefs in the new paradigm (Lewin 1952). This is summarized in the table below.

Stage of change in paradigm	Type of experiment
Existing paradigm has problems dealing with anomalies (unfreeze)	Destructive thought experiment
Attempts are made to find new paradigms (move)	Elucidative thought experiment
	Conjectural thought experiment
New paradigm is found and becomes dominant (refreeze)	Conclusive thought experiment

2.3 Scientific Revolution in Information Systems?

It is our conjecture that the information systems discipline is now in a Kuhnian crisis. The gap between the expectations raised by the functional/representational paradigm (traditionally at the heart of the discipline) and the growing number of anomalies that "do not respond to minor adjustments of the existing (functional/representational) conceptual and instrumental fabric" (Kuhn 1977, p. 262) is bigger than ever.

There are already signs that researchers are critically analyzing the fabric of their beliefs that have brought the community to its current position. There are papers and conference panels that reconsider the fundamental assumptions behind the way that information systems research is currently being done (Hirschheim, Klein and Lyytinen 1996), on whether information systems can survive as a unique discipline (Markus 1996) (Fedorowicz et al. 1996), on the way that information systems uses research from other disciplines (Adam and Fitzgerald 1996), on the relative merits of different research approaches (Newsted et al. 1996; Myers et al. 1996) and on the effectiveness of current techniques (Fitzgerald 1996).

Kuhn states that during times of crisis "even philosophy will become a legitimate scientific tool" (p. 263) . The 1996 AIS Americas conference had twenty-one papers on the philosophical foundations of information systems. The previous year there had only been a (poorly attended) panel.

Thus there is growing evidence to suggest parts of the discipline are becoming uneasy with the beliefs and assumptions that underlie the dominant paradigm of information systems. In this atmosphere, it is timely to reintroduce the notion of thought experiments to the information systems community as a legitimate method for doing qualitative research and for advancing the field of information systems. Using our framework, thought experiments can help provide the mechanism for making the shift between dominant modes of thought.

3 AN ANALYSIS OF SOME THOUGHT EXPERIMENTS

This section analyzes four thought experiments that are related to information systems research. Each experiment is described and its impact on the field is outlined. Finally, an attempt is made to classify the experiment in the taxonomy given above.

3.1   Searle's Chinese Room

This thought experiment was presented by the philosopher John Searle to examine the notion that computers can understand language (Searle 1980). Searle asks us to *imagine* a room with two slots in the door. People walk up to the room and place two sheets of paper with Chinese writing on them into the upper slot. Then, after a short period of time, they receive a third sheet of paper from the lower slot. Again this sheet has Chinese writing on it. More particularly, the first sheet of paper contains a story in Chinese, the second sheet is a series of questions about the story and the third sheet has the answers to the questions, all written in perfect Chinese. To the outside observer, therefore, the room would seem to understand Chinese. Inside the room, however, is Searle himself who, he assures us, cannot understand a word of Chinese. Instead, he has a series of charts and rule books that say things like "when you get these shapes on the papers respond by writing the following shapes on the sheet." Whenever he receives the story and questions he responds by writing the answers and returns them to the person outside.

The process described here is functionally equivalent to that given by Schank and Childers (1984) when describing their natural language understanding programs. Here the program is given a tale in natural language and a series of questions about the tale and the computer responds with answers to the questions. Schank and Childers claim, therefore, that the computer understands natural language. Not so, argues Searle, based on his thought experiment. Searle, sitting inside the room does not understand Chinese, he is simply manipulating symbols. It is only the external observer who attributes this understanding to the system (see Maturana and Varela

(1992) and Winograd and Flores (1986) for further examples of this difference between internal operation and attribution based on observation).

Searle's point with the thought experiment is to clarify the difference between simply manipulating symbols using rules and truly understanding language. Various attempts have been made to try and explain how this situation can lead to an understanding of language, often by explaining that it is not Searle who understands Chinese, but rather it is the combination of Searle, the instruction books and the room. Such explanations, however, do not really help clarify how manipulating symbols can lead to understanding.

Impact of the Chinese Room Thought Experiment. The Chinese Room experiment provoked many responses from supporters of artificial intelligence and its opponents (indeed, the initial publication of the paper was accompanied by a series of commentaries and a response by Searle). Perhaps one of the most interesting consequences is given by Turkle (1996), who states that the experiment suggested a way of dealing with the idea of intelligent machines.

> Searle diffused fears of a cybernetic Frankenstein by reassuring people that even advanced programs, far more complex than those that currently exist, do not embody intelligence in the way human beings do. So accepting machine intelligence became comfortable, because it could be seen as unthreatening to human uniqueness. [p. 124]

In this way, it is apparent that Searle's experiment fits within the destructive category in that it addresses an existing theoretical position and demonstrates that it is incompatible with the background beliefs about human understanding of natural language. It also highlights the existence of conceptual problems that are implicit in the existing view that have not, to date, been fully considered.

3.2 Introna's Transparent World

Introna (1996b) argues that one of the problems with privacy discourse is that it assumes we (society) know why we need privacy, or we know what we would lose if we lose privacy. Introna proposes a thought experiment to help clarify the impact of a loss of privacy on social relationships by proposing a world of absolute transparency.

Imagine a world where there is a comprehensive and complete lack of privacy, complete and immediate access, complete and immediate knowledge, and complete and constant observation of every individual. There will be no private thoughts and no private places. Every thought and every act is completely transparent from motive right through to the actual thought or behavior. Body and mind are immediately and completely transparent to each and every "other." Let us call this world the transparent world.

Let us consider the nature of social relationships in the transparent world with a number of questions. Would differentiated relationships be possible? How would your relationship with your wife or lover differ from your relationship with an official or your manager or your child? What will there be to exclusively *share* since

everything is always already known to every "other"? It seems that in the transparent world, notions such as getting to know someone, or being intimate with someone, or sharing yourself with someone just fade into obscurity.

Furthermore, in such a world, how would you differentiate yourself, how would you compete? Competitive advantage requires knowledge of a method, a technique or a way of doing that is not known to the other. Is creativity possible? How is it possible to say "this is my idea" or "this is what I think"? Does it make sense to talk of "my" or "me" at all, since original thought or original action would, in principle, be impossible (or at least indeterminable)?

From this thought experiment it is clear that all social relationships, relationships of collaboration or of competition, require at least some level of privacy.

Impact of Introna's transparent world. This is an example of a constructive conjectural thought experiment. In the experiment, one variable the level of transparency) is increased to its limit. This helps the researcher to think through the issues that may become gray somewhere in the middle. Now, one can protest against such a thought experiment by saying that the transparent world is a useless concept since total transparency is impossible. It is Introna's contention that the issues that are very apparent in the transparent world, as shown in the thought experiment, become issues in the everyday world, where technology erodes our experience or privacy, long before total transparency is reached and it is the thought experiment that allows us to focus on them.

3.3   Varela's Submarine

Cognitive science theory postulates that in order for a system (an individual) to act in a meaningful way in a world, the system must have access to some type of representation of that world. Hence, people have (make/remake) some form of representations of the world "out there" in their heads that functions as the "map" upon which their action is based. The actor would then become aware of certain environmental inputs and would use these inputs and the "map" to decide what the appropriate outputs should be.

Maturana and Varela (1992) argue that this theory is wrong. They argue that the problem is with our epistemological accounting. We are projecting the world of the scientist (as an observer of phenomena) onto the world of the involved actor, in the world. They argue that the involved actor does not take inputs and turn them into outputs. On the contrary, for the involved actor the environmental inputs are merely disturbances (perturbations) that threaten the internal coherence that needs to be maintained if the system is to survive.

To demonstrate this confusion they make use of a thought experiment of a submarine pilot (Maturana and Varela 1992, pp. 136–137). *Imagine* a person who *always* lived in a submarine. The person has never left the submarine but is trained how to pilot it. You (the observer) see the submarine float gracefully through very dangerous reefs and obstacles and surface in the still waters of the bay. On surfacing you contact the submarine pilot and announce: "Congratulations! You avoided the

reefs and surfaced beautifully. You really know how to pilot a submarine in dangerous circumstances." The pilot in the submarine is, however, perplexed: "What is this about reefs and surfacing? All I did was push some levers and knobs to establish certain relationships between indicators as I operated the levers and knobs. It was done in a prescribed sequence which I am used to doing. This is what my father taught me, and his father taught him. I did not do any special manoeuver, and on top of that, you talk to me about a submarine. What is that? You must be kidding!" All that exists for the pilot of the submarine are the different indicator readings, their transitions, and sequences of obtaining specific relations between them. It is only for the observer that the dynamic relations between the submarine and its environment exists. These are representations which the submarine pilot does not have or need.

In the example we can see that for the observer there is an input–output model of reality. The model described what actions (by the pilot) created what relationships between the submarine and its environment. For the pilot there is only a closure type description. The inputs from the environment were mere disturbances that had to be dealt with by restoring a particular relationship between certain indicators by applying certain levers or knobs. For the pilot there is an internal coherence that must be maintained. If this coherence is maintained, the submarine world will stay intact; if not, it will disappear.

Impact of Varela's submarine. The argument implicit in the submarine thought experiment has some very important implications for information systems research. If the central doctrine of representationalism used by information systems – imported from the reference discipline of cognitive science – is successfully challenged, then much of what is today seen as the core of the discipline would be under suspicion. For example, it can be used to argue that systems developers do not explicitly attend to method but use method only as part of "getting the job done" (Introna and Whitley 1996). It can also be used to argue that managers do not explicitly attend to information (as an input) but rather see it as perturbations that need to be mediated in "getting the job done" (Introna 1996a).

Varela's submarine is a clear example of a destructive thought experiment (against representationalism). It also serves as a constructive, conclusive thought experiment in support of the theory that meaningful action could be based on local structural coupling (Maturana and Varela 1992).

3.4 The Collins–Turing Test

The issue of intelligent machines is the focus of a second thought experiment, although this time the experiment is more concerned with clarifying misconceptions about *what it means to be a socialized human* than to prove the impossibility of artificial intelligence.

In 1950, the mathematician Alan Turing proposed a test that could be used to address the issue of whether or not a machine was intelligent. Instead of asking what tasks a machine would need to perform in order to be considered intelligent, Turing proposed a test that could be used to examine the claims that a machine was intelligent. Essentially, if a human interrogator could not differentiate between a machine and a human during a period of mediated interaction, the computer would be deemed to be intelligent.

The sociologist of knowledge, Harry Collins (1990), has proposed a thought experiment based on the Turing test which he claims shows the vital role that socialized knowledge plays in being human; knowledge which he argues cannot be explicitly taught to a computer.

When humans interact, they have an overwhelming tendency to compensate for their conversational partners. When people talk, they rarely do so using full, grammatically accurate sentences; their conversations are made up of pauses, mistakes and incomplete sentences. Nevertheless, most conversations do not break down at each of these problems. Instead the dialogue flows smoothly. To counter this tendency for compensation, Collins proposes that the experiment's protocol be tightened up to ensure that the judges are actively looking for the computer and will not make unnecessary allowances for the computer.

Collins' model of human knowledge argues that much of this knowledge is gained through socialization rather than explicit teaching. As such, he argues that unsocialized knowledge can only be spotted during long interactions which allow for the full possibility of interactions.

To demonstrate this, he proposes the following thought experiment. *Imagine* the range of interactions that can take place during a typical Turing test type interaction. Assume that the interaction lasts for one hour and, with both parties typing quickly, a total of 20,000 symbols can be created. If each symbol could be one of 100 possibilities, this gives a total 100 raised to the power of 20,000 possible strings of symbols ($10^{20,000}$). Now *imagine* (this is a thought experiment after all) that all these strings are stored in a computer. One would begin with "Hello how are you"; another would begin "Hello, how are yov," etc. The programmer of the computer now has to select which of these symbol strings to discard and which to keep for the intelligent machine. Many of these strings are simply not sensible (20,000 letter "A's," 20,000 letter "z's," 19,999 "A's" followed by a "B," etc.). However, others are sensible but would not necessarily be used by the programmer. The program works by waiting for the first statement and then picking, at random, one of the responses from the remaining options that match that part of the symbol string. Thus, if the interrogator begins by saying "Hello," the computer responds with one of its preprogrammed choices for following the initiation "Hello."

If the programmer deleted all the responses to a particular question, then the system would crash – it has no from which options to choose. It could, however, be programmed to degrade gracefully by apologizing for not understanding the question. A more difficult problem, however, arises with, for example, conversations that talk about politics. If the interrogator says "I prefer conservatives," what should the

programmer do, especially if the programmer's politics are left of center? One option would be to delete all the symbol strings that responded with "So do I" and the like.

By restricting such statements, however, the programmer is not accurately matching the true conversational repertoire of a socialized human. There are occasions when it would be sensible to make statements that give a different political opinion to the programmer's normal one. The canny interrogator would therefore simply have to look out for the conversational partner that never changed its political viewpoint, no matter what the circumstances.

There are further problems if the programmer chooses a more liberal policy. Consider the case of humorous statements and jokes. A more liberal approach would be to include jokes in the list of possible conversations. Unfortunately, a consequence of this would either be a partner who never made jokes (too few jokes were left in the system) or one who made too many jokes, again easily spotted.

The purpose of the Collins–Turing test is to demonstrate the important role that socialized knowledge (such as when it is appropriate to express a different political view or when a joke is called for) is something that cannot be programmed; it is not a rule, it is a learned skill about knowing when to break the rules of conversation.

Impact of the Collins–Turing test. The impact of the Collins–Turing test on the artificial intelligence community is interesting to observe. On the one hand, it is possible to argue that it has had a very limited impact because Turing tests are still being undertaken without taking into consideration the points made by Collins (Shieber 1994a, 1994b; Loebner 1994). On the other hand, Collins has had a far greater impact on the artificial intelligence community at large, most recently as the opening reviewer of the 1992 book by Hubert Dreyfus, *What Computers Still Can't Do,* for the leading journal *Artificial Intelligence*.

The conceptual clarity provided by Collins has been used to address the issue of whether or not it is possible to create false identities on the internet (Whitley 1996). Much has been written about the possibility of creating false identities in cyberspace, where physical cues are no longer present; all that the conversational partner has access to is the linguistic utterances of the speaker. In such a case, however, the person attempting to create a false identity has exactly the same problems as the programmer in the thought experiment. What kind of utterances should the fake character make? In the long run, it will not be possible for the fakers always to make the appropriate utterances because they haven't been socialized into that role. Thus, while it is possible to appear to be a false identity, in the long term this false identity will be spotted, and it is most likely to be spotted by someone who has been socialized into the identity being mimicked.

This is a constructive, conclusive thought experiment in that it proposes a new way of viewing human intelligence in terms of socialized knowledge. As with Searle's experiment, it does not just propose a hypothetical situation but develops the implications of the situation in a variety of ways.

4 ENSURING EFFECTIVE THOUGHT EXPERIMENTS

As the previous examples have shown, in order for a thought experiment to be effective, it must work from within the existing conceptual apparatus of researchers. That is, it must be based on uncontested beliefs (dogmas) of the field. In particular, this means that the experiment must be based on known and accepted terms and concepts, it must articulate all assumptions needed as part of the narrative and these assumptions must also be reasonable to the field as it is. Thus the narrative must ensure that the impact of the thought experiment is not deflected by issues of implausibility (Boland and Schultze 1995). "The imagined situation must allow the researcher to employ his usual concepts in the way he has employed them before" (Kuhn 1977, p. 265).

   Based on known and accepted terms and concepts. The terms and concepts employed in the thought experiment must be generally accepted within the current paradigm or must be such that the normal everyday understanding of the terms or concepts will suffice. If this is not true, the revolutionary power of the thought experiment is lost. In Introna's transparent world, the notion of sharing in interpersonal relationships may be too vague and ambiguous to make the experiment convincing, although one of the intentions of the experiment is to help focus on what is meant by sharing in precisely such situations. In contrast, Searle makes every effort to ensure that the "Searle" in the Chinese room acts according to known concepts of computing, and in particular mimics the processes of the computer systems that, it is claimed, can understand natural language.

   Background must be specified. As with other experiments, thought experiments should also take care to make explicit background assumptions part of the narrative. Thus, in Varela's experiment, the narrative has the observer speaking effortlessly to the pilot of the submarine even though the pilot does not, by the terms of the experiment, share the life world of the person on the shore (Habermas 1984). Some of the impact of the experiment, about the different epistemological boundaries, is lost as one tries to understand how they could possibly talk together, given their vastly different experiences. If Varela had included an explicit background assumption about their ability to communicate, this problem would not be raised.

   The purpose of the thought experiment is not simply to result in a further refinement of the field to cope with yet another anomaly. It is intended to encourage a shift in paradigm. Therefore the full rhetorical tools of discourse must be used to their best effect (Fuchs 1992, p. 30). Unlike traditional experiments, whose results are always under-determined by the data, thought experiments can't be explained away by poor experimental practice or misreading. Either the thought experiment succeeds or it doesn't; it fails in terms of its rhetoric, not its implementation.

5 CONCLUDING DISCUSSION

In this paper, we introduced the notion of a thought experiment as a basis for undertaking qualitative research. It was argued that thought experiments have a particularly important role to play in terms of Kuhnian scientific revolutions.

Next, we sought to classify thought experiments in terms of whether the results they produce are destructive or constructive. A destructive thought experiment seeks to show that existing ways of viewing the world contain conceptual problems that the experiment elicits. Constructive experiments offer new ways of viewing the world.

We then described four thought experiments that could affect our understanding of computer-based systems. These were Searle's Chinese room, Introna's transparent world, Varela's submarine, and Collins' revised Turing test. Each experiment was described and its impact assessed. The experiments were chosen to illustrate the range of uses to which thought experiments have been put. Searle's experiment describes a hypothetical situation and uses it to destroy the belief that simply manipulating symbols is the same as understanding language. Introna, in contrast, offers a more constructive experiment in which the hypothetical situation is used to clarify what is meant by shared relationships.

Thought experiments are effective only rhetorically. It is therefore essential that their rhetorical effect is maximized; in the paper, we reviewed the conditions necessary to achieve this. In particular, the notion that the thought experiment, while proposing a new world view has to be understandable and acceptable from the framework of the existing belief systems.

We have shown how thought experiments can be used to enable paradigm shifts. The information systems discipline, at the current time, seems to be showing symptoms of a Kuhnian crisis. It is becoming clear that the dominant paradigm, based on rationalist and functionalist beliefs, and the main research methods used, namely hypothesis testing and qualitative data analysis, are inappropriate for the full range of phenomena associated with computer-based systems in organizations. Moreover, an alternative paradigm, loosely based around interpretivism (Walsham 1993, 1995), exists with a growing following. It is in this context that we propose that the information systems discipline look more closely at thought experiments as a means of clarifying its conceptual apparatus; of deciding whether it wishes to irreversibly to change its foundational beliefs. If so, we look forward to more published thought experiments in the field to encourage scientific disruption. Perhaps advocates of interpretivism, or any other emerging research approach, should attempt to construct convincing thought experiments to encourage us to follow their approach.

6 ACKNOWLEDGMENTS

7 REFERENCES

Adam, F., and Fitzgerald, B. (1996) "A Framework for Analyzing the Evolution of the IS Field: Can IS Become a Stable Discipline." In J. D. Coelho, T. Jelassi, W. König, H. Krcmar, R. O'Callaghan, and M. Sääksjarvi (Editors), *Fourth European Conference on Information Systems.* Lisbon, Portugal, pp. 17-32.

Boland, R. J., and Schultze, U. (1995). "From Work to Activity: Technology and the Narrative of Progress." In W. J. Orlikowski, G. Walsham, M. R. Jones, and J. I. DeGross (Editors), *Information Technology and Changes in Organizational Work.* London: Chapman and Hall, pp. 308-324.

Brown, J. R. (1986). "Thought Experiments Since the Scientific Revolution." *International Studies in the Philosophy of Science*, Volume 1, Number 1, pp. 1-15.

Bunzl, M. (1996). "The Logic of Thought Experiments." *Syntheses,* Volume 106, Number 2, pp. 227-240.

Collins, H. M. (1990). *Artificial Experts: Social Knowledge and Intelligent Machines.* Cambridge, Massachusetts: The MIT Press.

Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason.* Cambridge, Massachusetts: The MIT Press.

Einstein, A. (1949). "Autobiographical Notes." In A. Schlipp (Editor), *Albert Einstein: Philosopher-Scientist*. LaSalle, Illinois: Open Court.

Fedorowicz, J.; Lee, D.; Sviokla, J. J.; and Weber, B. (1996). "Are There Barbarians at the Gates of Information Systems?" Panel presented at the Seventeenth International Conference on Information Systems, Cleveland, Ohio, December 16-18.

Fitzgerald, B. (1996) "An Investigation of the Use of System Development Methodologies in Practice." In J. D. Coelho, T. Jelassi, W. König, H. Krcmar, R. O'Callaghan, and M. Sääksjarvi (Editors), *Fourth European Conference on Information Systems*. Lisbon, Portugal, pp. 143-161.

Fuchs, S. (1992). *The Professional Quest for Truth: A Social Theory of Science and Knowledge.* Albany: State University of New York Press.

Habermas, J. R. (1984) . *The Theory of Communicative Action.* London: Heinemann Education.

Hirschheim, R.; Klein, H.; and Lyytinen, K. (1996). "Exploring the Conceptual Foundations of Information Systems." *Accounting, Management and Information Technology.* Forthcoming.

Introna, L. D. (1996a) *Management, Information and Power.* London: Macmillan, forthcoming.

Introna, L. D. (1996b). "Privacy and the Computer: Why We Need Privacy in the Information Society." London School of Economics.

Introna, L. D., and Whitley, E. A. (1996). "Against method." London School of Economics.

Kuhn, T. S. (1977). "A Function for Thought Experiments." In T. S. Kuhn (Editor), *The Essential Tension: Selected Studies in Scientific Tradition and Change.* Chicago: University of Chicago Press, pp. 240-265.

Lewin, K. (1952). "Group Decision and Social Change." In T. Newcomb, E. Hartley, and E. MacCobby (Editors), *Readings in Social Psychology.* New York: Holt, Rinehart & Winston.

Loebner, H. G. (1994). "In Response." *Communications of the ACM*, Volume 37, Number 6, pp. 79-82.

Markus, M. L. (1996). "Thinking the Unthinkable: What Happens if the IS Field as We Know It Goes Away?" Claremont Graduate School.

Maturana, H. R., and Varela, F. J. (1992). *The Tree of Knowledge: The Biological Roots of Human Understanding.* Revised edition. Boston: Shambhala.

Myers, M. D.; Klein, H. K.; Truex, D.; and Wynn, E. (1996). "The Merits of Three Qualitative Research Methods." Panel presented at the Seventeenth International Conference on Information Systems, Cleveland, Ohio, December 16-18.

Newsted, P.; Chin, W.; Ngwenyama, O. K.; and Lee, A. (1996). "Resolved: Surveys Have Outlived Their Usefulness in Information Systems Research." Panel presented at the Seventeenth International Conference on Information Systems, Cleveland, Ohio, December 16-19.

Schank, R. C., and Childers, P. G. (1984). *The Cognitive Computer: On Language, Learning, and Artificial Intelligence.* Reading, Massachusetts: Addison-Wesley Publishing Company.

Searle, J. R. (1980). "Minds, Brains and Programs (with Commentary and Author's Response)." *The Behavioral and Brain Sciences* , Volume 3, Number 3, pp. 417-457.

Shieber, S. M. (1994a). "Lessons from a Restricted Turing Test." *Communications of the ACM*, Volume 37, Number 6, pp. 70-78.

Shieber, S. M. (1994b). Your Article on the Turing Test, June 1994. Personal communication.

Sorensen, R. A. (1992). "Thought Experiments and the Epistemology of Laws." *Canadian Journal of Philosophy*, Volume 22, Number 1, pp. 15-44.

Turkle, S. (1996). *Life on the Screen: Identity in the Age of the Internet.* London: Weidenfeld & Nicholson.

Walsham, G. (1993). *Interpreting Information Systems in Oganisations.* Chichester, England: John Wiley.

Walsham, G. (1995). "The Emergence of Interpretivism in IS Research." *Information Systems Research*, Volume 6, Number 4, pp. 376-394.

Whitley, E. A. (1996). "Is It Possible to Really Play with Identity in Cyber'space'? A Review Based on the Sociology of Knowledge." London School of Economics.

Wilkes, K. V. (1988). *Real people: Personal Identity without Thought Experiments.* Oxford: Clarendon Press.

Winograd, T., and Flores, F. (1986). *Understanding Computers and Cognition: A New Foundation for Design.* Reading, Massachusetts: Addison Wesley.

Wittgenstein, L. (1969). *On Certainty.* Denis Paul and G. E. M. Anscombe, Translators. Oxford: Basil Blackwell.

8 BIOGRAPHY

Lucas D. Introna is a Lecturer in Information Systems at the London School of Economics and Political Science and Visiting Professor of Information Systems at the University of Pretoria. His research interest is the social dimensions and consequences of information technology. He is associate editor of *Information Technology & People* and *The South African Computer Journal*. His practical and consulting experience includes design and implementation of information systems for various large corporations as well as strategic management consulting.

Edgar A. Whitley is a lecturer in Information Systems at the London School of Economics and Political Science. He has a B.Sc. (Econ) Computing and a Ph.D. in Information Systems, both from the LSE. He has taught undergraduate, postgraduate students and managers in the UK and abroad. Edgar was one of the organizers of the First European Conference on Information Systems and is actively involved in the coordination of ECIS conferences. He has published widely on various information systems issues and is currently completing a book on the socio-philosophical foundations of information systems.