SPARC 2006

Proceedings of the Salford Postgraduate Annual Research Conference

11-12 May 2006

Table of Contents

Preface

It gives me great pleasure to introduce this collection of papers from the 2006 Salford Postgraduate Annual Research Conference (SPARC). This conference represents an ideal opportunity for postgraduate researchers to present their work in a friendly yet rigorous environment, as a preparation for taking that work further afield in international conferences. The quality of the papers and posters at the conference, as the contributions in this collection demonstrate, is always of the highest standard. These papers only represent a fraction of those presented at SPARC, but their diversity and strength epitomises the invaluable contribution that the research community represents for the University. As this collection demonstrates, the wide variety of intellectual disciplines in which Salford researchers undertake their research is a testament to the thriving environment which they both enjoy and contribute to.

These proceedings also represent a milestone for SPARC itself. The 2006 conference will have been the last 'closed' event, with SPARC 2007 opening its doors to postgraduate researchers from institutions in the North-West and beyond. What started in the late 1990s as a relatively modest gathering has now grown to include over 100 participants. We want to build on this momentum to bring SPARC to an ever-wider audience so that postgraduate researchers beyond Salford can benefit from participation. Equally, by opening the conference to external presenters, the quality, coherence and academic value can only increase. Moreover, if the international conference circuit represents as much a forum for networking as for academic presentation, it would be remiss not to introduce this element to our postgraduate community.

Lastly, that the conference has continued to be such a success is a tribute to the hard work of the organising team as well as the participating academics. I would like to thank Sachin Anand and Linda Kelly for their outstanding work in this regard, and also Giles Simon who has done an excellent job of copy-editing the eclectic array of papers in this collection.

Dr Jocelyn AJ Evans
Director of Graduate Studies
University of Salford

Modelling Hotel Service Failure and Recovery from the Consumer Perspective

Ian D Bennett

Management and Management Sciences Research Institute

Introduction

Hotels are characterised by continuous operation, highly fluctuating demand relative to constant rates of supply and a high degree of interaction between employees and consumers; consequently few organisations can deliver 'zero defects' service every time. Not surprisingly, the management of complaints about service failure is becoming increasingly important not least because it has a significant impact on consumer satisfaction (Kelley et al. 1993), customer confidence (Boshoff 1997; Boshoff and Leong 1998), word-of-mouth recommendation (Bailey 1994; Mattila 2001), intention to repurchase (Morris 1988; Smith and Bolton 1998), costs and revenue (Armistead et al. 1995). Davidow (2000) argues that service recovery is successful only to the extent that a complainant continues to repurchase the service and engages in positive word-of-mouth communication about his/her experience. That being the case, it is critical that researchers and managers understand the way in which consumers react to service failure and recovery, and the interrelationship between them with respect to the salience of service attributes, the effectiveness of recovery strategies employed and concepts of blame attribution (Bitner 1990; Weiner 1982; Hewstone 1989) and perceived justice (Tyler 1994; Sparks and Callan 1996).

This study is concerned with service failure and recovery strategies within the context of service quality and consumer behaviour in the hotel sector. The research is different from previous empirical work in this subject area in three important ways. Firstly, in contrast to both the large majority of published research on service quality and all previous empirical work on service failure and recovery, the conceptual framework is not based on the expectancy-(dis)confirmation paradigm (where consumers evaluate a service by comparing pre-consumption *expectations* with actual *performance*). Instead, the research problem has been contextualised within a process-based model using service quality *importance* and *performance* constructs as predictors of consumer satisfaction and loyalty (Figure 1).

Secondly, the analysis of service failure and recovery is embedded in the context of the consumers' perceived importance of service attributes (Figure 2). It is considered that the perceived importance of service attributes that fail *ceteris paribus* will significantly influence consumer reaction to the failure, the perceived effectiveness of recovery strategies and, in turn, the outcome of critical incidents with respect to overall satisfaction and loyalty. Therefore, whilst the 'performance-only' approach is now generally regarded as the most effective model (Churchill and Surprenant 1982; Carman 1990; Cronin and Taylor 1992) in terms of its superior predictive validity over 'importance-performance' models (performance weighted by importance) in studies of consumer satisfaction (*inter alia* Churchill and Surprenant 1982; Cronin and Taylor 1992; Crompton and Love 1995; Yuksel and Rimmington 1998), in the particular context of service failure and recovery, it is expected that the 'performance weighted by importance' model will have greater predictive validity. Ostrom and

Iacobussi (1995) and Hoffman and Kelly (2000) have shown that the evaluation of a service encounter is influenced by the perceived importance or criticality of the outcome. In situations of high criticality, consumer involvement increases, and given the importance-performance response function, i.e. an inverse relationship between attribute importance and performance as consumer needs are met (Duke and Persia 1996; Sampson and Showalter 1999), it is possible that when a service fails and needs are not met, the perceived importance (criticality) of that service increases. Moreover, the use of the *importance* construct will increase the diagnostic ability of the theoretical model with respect to both service failure and service recovery elements of the study; research in this area has hitherto examined service failure and recovery without considering attribute salience.

Thirdly, the study will focus on three and four star hotels[1] in the UK to identify the relative importance of relevant service attributes, the tangible, functional and environmental dimensions of service failure and a range of recovery strategies in order to examine the key concepts of 'distributive', 'interactional' and 'procedural' justice from the consumer perspective and their effects on satisfaction and loyalty. The concept of 'perceived justice' refers to the fairness with which individuals have been treated by an organisation (Campbell and Finch 2004). The two fundamental building blocks of organisational justice theory are 'distributive' and 'procedural' justice. Distributive justice refers to the perceived fairness of the outcomes that an individual receives from organisations (Folger and Cropanzano 1998). However, given that the procedures used to determine outcomes can be more influential than the outcome itself, the emphasis has gradually shifted from distributive to procedural justice. This is defined as the perceived fairness of the procedures used in making decisions (Folger and Greenberg 1985). A third dimension relating to the latter is 'interactional' justice, which is concerned with the sensitivity with which information is communicated by employees to guests.

Within the proposed framework, there are a number of specific objectives. These are either based on recommendations from previous research or on identified gaps in the literature.

- Identify the relative importance of hospitality service attributes;
- Identify and classify service failure attributes and dimensions as perceived by customers in 3 and 4 star hotels;
- Identify and classify the service recovery strategies used by 3 and 4 star hotels as revealed from real-life experiences of customers;
- Test the theoretical model and evaluate the role and predictive validity of 'importance', 'performance' and 'importance-performance' constructs in the context of service failure, recovery, satisfaction and loyalty;
- Evaluate the effect of 'direct' and 'indirect' measures of service attribute importance on service failure, recovery, satisfaction and loyalty;
- Determine the relative effectiveness of service recovery attributes and dimensions in relation to service failure, satisfaction and loyalty;

[1] Three and four star hotels were selected because they represent organisations that provide high quality service on a range of attributes and invite high expectations from the consumer; service failure was therefore considered to be more likely than in five star hotels and more critical to consumer satisfaction and subsequent behaviour than in 1 and 2 star hotels.

- Determine the relative effectiveness of 'technical' (core), 'functional' (relationship) and 'environmental' dimensions in relation to service failure, satisfaction and loyalty;
- Analyse the effectiveness of service recovery strategy attributes and dimensions in relation to service component importance and severity of failure;
- Evaluate the strengths and weaknesses of strategy options with respect to a range of service failure scenarios;
- Determine the critical factors in distinguishing between consumers in relation to perceived service quality attribute importance, service failure and recovery evaluation, satisfaction and loyalty.

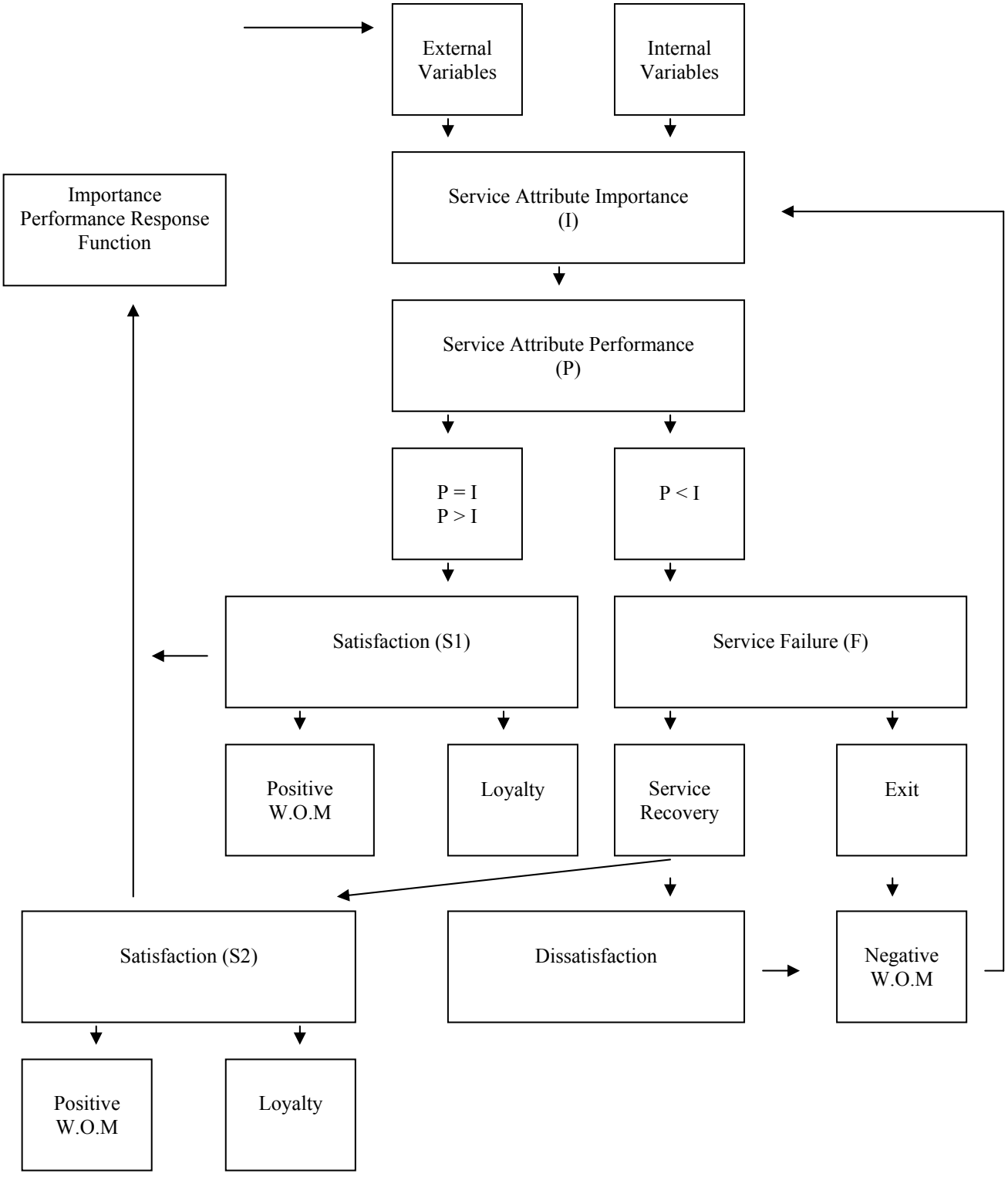
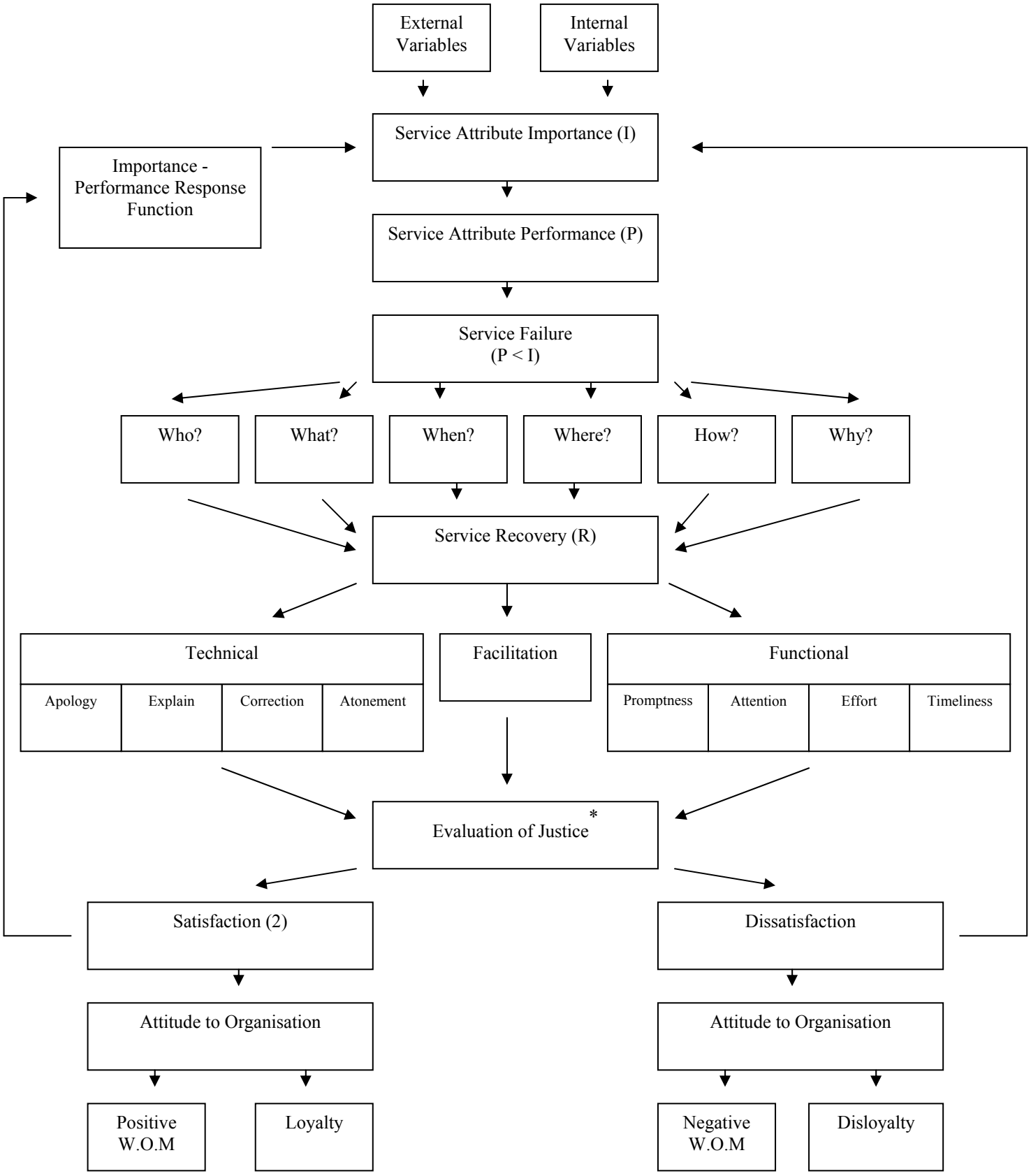


Figure 1: Service Importance, Performance, Failure and Recovery



* Distributive justice; interactional justice; procedural justice

Figure 2: Service Failure and Recovery

Literature Review

Theoretical Paradigms

Much of the research hitherto undertaken in service quality has been based on the expectancy-(dis)confirmation paradigm which proposes that consumers evaluate a product/experience by comparing pre-consumption *expectations* with the perceived *performance* of its attributes. This technique has been operationalised in the SERVQUAL model (Zeithaml et al. 1985), which has been used extensively in a wide range of service quality studies including hospitality research. Despite its continued use in both service quality and service failure and recovery research, the expectancy-disconfirmation paradigm has received considerable theoretical and operational criticism in the literature. Firstly, the use and interpretation of discrepancy scores has been questioned because it is the researcher who performs the comparison between expectations and performance perceptions (Crompton and Love 1995; Oh 1999). The validity of the 'difference' score as an indicator of quality or satisfaction has also been questioned because consumers may be satisfied even though there is negative disconfirmation (Hughes 1991; Pearce 1991). Secondly, there is consensus that the conceptualisation of expectations as a comparative standard is vague (Ekinci et al. 2000); they may or may not be based on experience (Carman, 1990) and/or could be 'ideal', 'predicted', 'minimum tolerable' or 'deserved' expectations (Miller 1977) or a combination of some or all of these, thereby affecting the direction and extent of the (dis)confirmation. A third criticism of this paradigm focuses on the timing of expectations measurement (Crompton and Love 1995; Yuksel and Rimmington 1998). If expectations are not solicited *a priori*, they are contaminated through experience (Carman 1990; Getty and Thomson 1994). Finally, there is no consensus on what the 'difference' score between expectations and performance actually represents i.e. is it a *service quality* or a *satisfaction* measure?

By comparison, the 'performance-only' approach to service quality and consumer satisfaction proposes that evaluations of a product/experience are affected by perceptions of the *performance* alone and that this is a more effective measure or indicator of quality and/or satisfaction (Churchill and Surprenant 1982; Cronin and Taylor 1992). Performance-only measures are more typical of the cognitive process (Meyer and Westerbarkey 1996) and play a pre-eminent role in the formation of customer satisfaction because performance is the main feature of the consumption experience (Yuksel and Rimmington 1998). Moreover, satisfaction results if a product 'performs' well irrespective of any disconfirmation effect (Mannell 1989). A number of studies have weighted service product attribute *performance* with *importance* scores to either determine perceived strengths and weaknesses from the consumer perspective (Mill 1989; Geva and Goldman 1991; Kozak and Nield 1998; Schofield, 2001) or to compare the validity of the different conceptualisations of quality and satisfaction as derivatives of *expectations, importance* and *performance* constructs (Dorfman, 1979; Fick and Ritchie; 1991; Crompton and Love 1995; Yuksel and Rimmington, 1998; Fallon and Schofield 2004). In all cases, the performance-only approach emerged as the most valid measure of quality/satisfaction, the disconfirmation-based models were the least valid and importance-weighting did not improve the predictive power of the measures. Despite these findings, widespread support exists for incorporating the importance construct into customer evaluation studies (Crompton and Love 1995; Oh and Parks 1998; Yuksel and Rimmington 1998; and Crompton 1999) because it provides a useful context and diagnostic tool,

which yields both insights about product attributes from the perspective of different consumers (Litvin and Ling 2001) and opportunities for effective resource management (Martilla and James 1977; Huan, Beaman and Shelby 2002; Khan 2004).

Service Failure and Recovery[2]

Service failures are defined as any perceived service related problems that transpire during a customer's experience with a firm (Maxham 2001). Service failure has been categorised by Bitner et al. (1990) according to employee behaviour relating to: the core service; requests for customised service; and unexpected employee actions. A later study (Bitner et al. 1994) included a typology of problematic customers, while Kelley et al. (1993) and Hoffman et al. (1995) added product and policy failures respectively. Johnston (1994) classified sources of failure on the basis of whether they were attributable to the organisation or to the customer and similarly, Armistead et al. (1995) offered 'customer error' and 'service provider error' with an additional 'associated organisational error'.

Customers respond to service failure in a number of ways. They may feel negative emotions - anger, disgust or contempt - based on their attributions about who is to blame (Smith and Ellsworth 1985; Godwin et al. 1995). Day and Landon (1976) argue that they first decide whether to convey an expression of dissatisfaction (action) or take no action; then, they will decide whether the response taken is public or private. Day (1984) examined complaint behaviour from the perspective of the goals being sought and identified 'redress seeking', 'complaining' and 'personal boycotting' behaviours. Similarly, Singh (1988) identified three sets of responses: 'voice responses', 'private responses' and 'third party responses'. This was later revised to four complaint response groups: 'passives', 'voicers', 'irates' and 'activists' (Singh 1990). Dissatisfied customers are more likely to complain to companies with a reputation for being responsive (Day and Landon 1977; Ganbois et al. 1977) or if a service is complex, expensive, important or if failure is serious (Blodgett and Granbois 1992; Bolfing 1989). Moreover, complaints are more likely to come from higher socio-economic groups (Singh 1990; Moyer 1984), from people who believe complaining will make a difference (Blodget and Granbois 1992; Day and Ash 1979) and from assertive people (Bolfing 1989; Fornell and Westbrook 1979).

Lewis and McCann's (2004) study of service failure in the UK hotel industry identified 26 types of service failure problems, the most common of which (experienced by 61.7 per cent), was slow restaurant service, followed by inefficient staff for business guests (60.3 per cent) and slow check-in/out for leisure guests (50.5 per cent). The respondents then considered the seriousness of each of the 26 service problems from the perspective of both leisure and business guests. Overall, the most serious failure was the room not being clean, followed by missing reservation, unfriendly and unhelpful staff, other staffing issues, poor locks, poor quality food and beverages, and an incorrect bill. Of the ten most serious failures, five related to

[2] An extensive review of the pertinent literature has been completed to inform both the theoretical framework and the proposed methodology; a selection of the more influential material is presented here.

staffing issues, suggesting the importance of the attitudes and conduct of staff in hotels.

Service recovery is the process by which a firm attempts to rectify a service delivery failure (Kelley and Davis 1994). Armistead et al. (1993), defined service recovery as specifications required to ensure that the customer receives a reasonable level of service after problems have occurred to disrupt normal service. By comparison, Johnston (1994) defines service recovery as seeking out and dealing with service failures as distinct from complaint handling, as many dissatisfied customers do not complain (Smith et al. 1999); it involves actions designed to resolve problems, alter negative attitudes of dissatisfied customers and to ultimately retain these customers (Miller et al. 2000). Successful service recovery has significant benefits, (Lewis and McCann 2004), it can enhance customers' perceptions of the quality of the service as well as the organisation, it can also lead to positive word-of-mouth behaviour as well as building customer relationships and maintaining loyalty (Bitner et al. 1990). Zemke and Bell (1990) attempted to define the content of service recovery, and concluded that customer expectations for service recovery are:

- to receive an apology for the fact that the customer is inconvenienced;
- to be offered a fair fix for the problem;
- to be treated in a way that suggests the company cares about fixing the problem and about the customer's inconvenience; and
- to be offered value-added atonement for the inconvenience.

Zemke and Bell (1990) outlined a number of steps required for effective service recovery; they suggested apology, re-instatement, empathy, symbolic atonement or compensation, and follow-up. Armistead et al. (1993) also suggested the importance of recovery follow-up activities and emphasised the need for speed of response. Clark et al. (1992) and Adamson (1993) also found that a speedy response, if only a written communication in the form of a letter, has a positive effect on consumer perceptions. Hart et al. (1990) argued that it was beneficial for companies to recover the situation, even if the cause of the service failure was the customer's fault because of the word-of-mouth benefits involved.

McDougall and Levesque's (1999) study of the effectiveness of service recovery strategies, in a range of sectors, showed that apology only, assistance, compensation and assistance plus compensation did not lead to positive future intentions towards the service firm. While assistance plus compensation was the most effective strategy, respondents still held negative future intentions towards the service firm, indicating that that current industry recovery practices were inadequate in mitigating negative intentions and that considerable effort to overcome customers' negative intentions is required.

The importance or criticality of the service has been found to effect consumer evaluations of the service outcome (Ostrom and Iacobucci 1995; Hoffman and Kelly 2000). In situations of high criticality, consumer involvement increases as does their desire for excellent service. As a result, the perceived importance of a service is likely to influence the consumer's reaction to service failure and the magnitude of the consequences associated with service failure is likely to increase as the importance of the purchase occasion intensifies. Similarly, in high criticality situations, service

recovery strategies that offer anything other than correcting the failed situation are perceived to be sub-standard. It is important to note that criticality relates to the service overall rather than particular attributes of a service. In the current project, the importance of *individual* service attributes and components will be assessed rather than the overall importance of the service.

Davidow (2000) argues that complainants evaluate organisational responses to service failure on the basis of six dimensions: apology, timeliness (speed of response to a complaint), facilitation (policies, procedures and tools to support responses to complaints), redress (the actual outcome received from the organisation), credibility (the organisation's willingness to account for the problem) and attentiveness (the interaction between staff and customer). Karatepe and Ekiz (2004) recognise seven similar dimensions: apology, promptness, facilitation, atonement, attentiveness, explanation (for the service failure) and effort (how hard staff try to resolve the problem and/or restore equilibrium).

Jenks (1993) argues that customers expect an apology and should always be given one. However, Zemke (1994) argues that leas than 50% of organisational complaint responses include an apology. Arguably an apology is the customers' 'psychological compensation' in that it assists them to restore equilibrium (Davidow 2000). It is not necessarily an admission of guilt on the part of the organisation, rather an indication that the problem has been taken seriously (Goodman, Malech and Boyd 1987). Moreover, Davidow (2000) argues that an apology is important because there is no financial cost involved and it significantly increases positive word-of-mouth publicity. Research by Martin (1985) and Conlon and Murray (1996) showed that response speed is positively related to satisfaction, repurchase intentions and decreased negative word-of-mouth publicity. Facilitation has a positive relationship with satisfaction with respect to complaint handling (Goodwin and Ross 1992). In other words, an organisation has a better chance of retaining a customer by encouraging them to complain, and then addressing that complaint, than it does by assuming that non-complaining customers are satisfied (Spreng, Harrell and Mackoy 1995). Indeed, Blodgett, Wakefield and Barnes (1995) found that facilitation decreased negative word-of-mouth.

An organisation's perceived credibility in terms of accounting for or explaining the reasons for a service failure and what will be done to prevent future repetition can influence customer satisfaction, repurchase intentions (Morris 1988) and word-of-mouth activity (Lewis 1983). The critical role of communication between the organisation and the customer is also reflected in the importance of attentiveness in complaint management. The interaction between company representatives and the complainant can enhance or detract from consumer satisfaction (Garrett, Meyers and Carney 1991; Testa, Skaruppa, and Pietrzak 1998). This dimension includes respect and courtesy for the complainant (Bossone 1994; Martin and Smart 1994), empathy for the complainant's situation (Berry 1995; Zemke 1994), a willingness to listen (Plymire 1991; Whitely 1994). Goodwin and Ross (1989) reported that attentiveness has a positive effect on satisfaction with complaint handling and repurchase intentions. Lewis (1983) and Blodgett, Wakefield, and Barnes (1995) both showed that attentiveness could lower negative word-of-mouth activity as well as increase repurchase intentions. Davidow (2000) also found that attentiveness and credibility had a major impact on complainant satisfaction, repurchase intention and the

likelihood of engaging in word-of-mouth activity. Effort, the seventh dimension identified by Karatepe and Ekiz (2004), is also important in that it makes a strong statement to the customer that the representative wants to help (Bell and Zemke 1992; Bossone 1994). The importance 'core' dimensions notwithstanding, previous research has highlighted the significance of recovery personnel's interpersonal skills in consumer satisfaction and future behaviour.

Under certain circumstances there may even be a 'recovery paradox' (McCollough and Bharadwaj 1992), i.e. customers may rate an organisation's performance higher if a delivery failure occurs and the organisation and its employees recover from the failure, than if the initial problem does not occur (Oliver 1997). The relationship between service recovery, satisfaction and behavioural intention has been addressed earlier. That between satisfaction and intention to repurchase has been supported by empirical research (*inter alia* Conlon and Murray 1996; Smith and Bolton 1998). The relationship between satisfaction and word-of-mouth activity is more complex in that satisfaction is negatively related to word-of-mouth likelihood (Lewis 1983; Walsh 1996) and positively related to word of mouth valence with complainers tending to behave consistently with the word-of-mouth that they give (Tax and Chandrashekaran 1992). Davidow (2000) found that response speed and the redress offered to complainants affected consumer satisfaction and the content of word-of-mouth activity, i.e. they had an indirect effect on future sales. Moreover, his results showed that whilst attentiveness may affect the likelihood of word-of-mouth activity, it is timeliness, apology and credibility that drive its content, which suggest that positive word-of-mouth activity would benefit from investment in these areas. Researchers have found a number of dimensions relating to consumer intention to repurchase products. Lewis (1983), Fornell and Wernerfelt (1988) and Spreng, Harrell and Mackoy (1995) found that redress had a direct positive effect on future sales (repurchase intention), whereas Davidow (2000) found that while redress had an important impact on satisfaction, it had no effect on repurchase intention.

Sparks and McColl-Kennedy (2001) explored a series of interactions between different aspects of perceived justice – extent of voice, level of concern displayed and extent of special treatment using a scenario-based approach. Their results highlight the importance of the way in which the service provider responds to customer complaints and the complexity of the interaction which affects consumer evaluations of service recovery.

Wirtz and Mattila's (2004) research on consumer responses to compensation, speed of recovery and apology after a service failure investigated how the three dimensions of distributive, procedural and interactional fairness influence consumers' attributional processes, their post-recovery satisfaction and behavioral responses (re-patronage intent and negative word-of-mouth communication). The results indicate that recovery outcomes (e.g. compensation), procedures (e.g. speed of recovery) and interactional treatment (e.g. apology) have a joint effect on post-recovery satisfaction. This suggests that compensation may not enhance satisfaction when the recovery process is well-executed (an immediate response combined with an apology). Similarly, compensation failed to lessen dissatisfaction with a poor recovery process (a delayed response without apology). However, offering compensation was effective in increasing satisfaction in mixed-bag recovery situations (delayed recovery with an apology, or immediate recovery without apology).

Mattila (2004) investigated the negative impact of service failure on customer loyalty with particular reference to the moderating role of affective commitment on post-failure attitudes and loyalty intentions under two service failure conditions: a successful and poor service recovery. The findings indicate that emotionally-bonded customers might feel 'betrayed' when a service failure occurs, resulting in a sharp decrease in post-recovery loyalty. Customers with lower levels of emotional bonding with the service provider were more 'forgiving' when the service recovery was effectively handled.

Weun, Beatty and Jones (2004) studied the impact of service failure severity, which has been previously linked to consumers' post-recovery perceptions (Smith and Bolton, 1998; Smith *et al.*, 1999) on service recovery evaluations and post-recovery relationships with respect to the 'core' aspects of service. They found that interactional and distributive justice were critical factors influencing customer satisfaction after a service recovery and that service failure severity had a significant influence on satisfaction, trust, commitment, and negative word-of-mouth publicity. They recommend that future research should seek to develop a more comprehensive model of service failure and recovery and that moderating factors such as service failure severity, consumer involvement with service and customer attitudes towards complaining should be investigated to assess their influence on customer perceptions of recovery strategies.

Research Hypotheses

A number of provisional research hypotheses have been developed from the literature review.

H1 Service failure on attributes of high importance has a significantly greater effect on post-recovery consumer satisfaction and loyalty than service failure on attributes of low importance.

H2 Where service failure occurs on attributes of high importance, consumer reaction to service recovery strategies is significantly different than for attributes of low importance.

H3 The 'importance-performance' (I-P) model explains more of the variance in post-recovery consumer satisfaction and loyalty than the performance-only' model.

H4 The 'indirect' I-P model explains more of the variance in post-recovery consumer satisfaction and loyalty than the 'direct' I-P model.

H5 The degree to which service attributes of high importance fail significantly influences the consumer's perception of the appropriateness of the recovery strategy employed.

H6 'Technical' response dimensions are more important than 'functional' response dimensions in effective service recovery (after Hurley 1998).

H7 'Apology' and/or 'explanation' are acceptable recovery strategies for service failure on attributes of low importance.

H8 'Assistance' (correction/recovery) and/or compensation are required for service failure on attributes of high importance.

H9 'Assistance' (correction/recovery) is significantly more effective then 'compensation' for producing higher levels of post-recovery consumer satisfaction and loyalty, when failure occurs on attributes of high importance.

H10 A recovery strategy involving 'apology' and 'assistance' is significantly more effective in producing consumer satisfaction and loyalty than either 'apology' or 'assistance' alone, when failure occurs on service attributes of high importance.
H11 'Compensation' is more effective than 'correction' for producing higher levels of post-recovery consumer satisfaction and loyalty, when failure occurs on attributes of low importance.
H12 Where service failure is severe on attributes of high importance, a recovery strategy that includes 'compensation' in addition to 'assistance' (correction/recovery) will produce higher levels of post-recovery consumer satisfaction and loyalty than a strategy one with 'assistance' only.
H13 Perceived 'distributive' justice is more important than 'interactive' justice, which in turn is more important than 'procedural' justice, in post-recovery consumer satisfaction and loyalty (after Hoffman and Kelly 2000).
H14 There is a significant interaction effect between perceived service attribute importance and the perceived effectiveness of the service recovery strategy with respect to post-recovery consumer satisfaction and loyalty.
H15 There are significant differences between leisure and business customers with regard to the perceived effectiveness of response dimensions when failure occurs on attributes of high importance.
H16 'Assistance' (correction/recovery) is the most effective 'technical' response dimension for leisure customers on attributes of high importance.
H17 'Compensation' is the most effective 'technical' response dimension for business customers on attributes of high importance.

Proposed Methodology

The study represents a cognitive-behavioural approach to the problem of service failure and recovery and will employ a post-positivist research design and a mixed-method approach to data collection, analysis and interpretation. Following accepted guidelines (Churchill 1979), the review of the pertinent literature will be followed by three stages of a primary research design. At stage one, qualitative research in the form of in-depth interviews with managers in three and four star hotels in Greater Manchester will be undertaken to discuss the following and thereby contextualise and operationalise theoretical concepts:

- key issues in achieving and maintaining service quality;
- relative frequency of technical, functional and environmental service failures;
- organisational recovery strategy options;
- effects of service failure on satisfaction and loyalty.

At stage two, a 'critical incident technique' or open-ended questionnaire survey will be used to elicit examples of service failure in three and four star hotels (Kelley et al. 1993; Johnston 1994; Hoffman et al. 1995; Tax et al. 1998; Miller et al. 2000; Lewis and Spyrakopoulos 2001; Lewis and McCann 2004). Subjects will be asked to describe in detail a prior service failure incident and the elements of their service failure experience including the cause of dissatisfaction, if and why they complained, who they spoke to together with the key response factors influencing their satisfaction with the response. This will be used to identify and classify both service failures and recovery strategies. The data will be content analysed and used to strengthen and validate the theoretical model. The strong precedent for scenario-based experiments

in service recovery research notwithstanding, following recommendations from *inter alia* Mattila (2004) subjects' direct experiences with service failures will be investigated (at stages two and three) because of their direct link with attitudes (Eagly and Chaiken 1993). This is considered to be more realistic than a simulated approach and more practical than considering the third option of investigating actual service failure in particular hotels.

At stage three, the theoretical framework and its constructs will be tested by being operationalised quantitatively, using multi-attribute Likert-type scales and comparative response formats (after Davidow 2000; Lewis and Spyrakopoulos 2001; Lewis and McCann 2004). A structured, self-complete questionnaire survey will be used. Subjects will be asked to complete a number of questions relating to service attribute importance, performance, satisfaction and future intention, and to rate service recovery strategies in relation to service failure, satisfaction and intention to both recommend and repurchase the service. A mix of multiple choice questions and balanced, seven-point Likert-type scales (anchored at *strongly disagree* (1) and *strongly agree* (7) with each intervening option being numbered and clearly labelled) will be used to elicit categorical and interval level data. An additional *don't know* option will also be used to differentiate between this category of respondents and those who *neither disagree/nor agree* (after Orams and Page 2000). Details relating to subjects' socio-demographic characteristics, personal characteristics and purchase motivation with respect to leisure or business will also be obtained to facilitate comparative analyses. Scenario-based approaches are also currently being considered in the event that subjects are unable to recall prior service failure incidents in sufficient detail during pilot tests of the instrument.

Summary

The use of a conceptual framework based on *importance* and *performance* constructs directly challenges the expectation-(dis)confirmation paradigm which still dominates research in service quality generally and in service failure and recovery in particular. Moreover, given the proven predictive validity of performance-only and importance-performance research designs, it is expected that the model will enable the researcher to make a significant contribution to the literature. Additionally, the diagnostic value of the *importance* construct including direct and indirect measures of service attribute importance and the in-depth analysis of service failure components, recovery strategy options, their interrelationship and effects on consumer satisfaction and loyalty, will present opportunities for further theoretical contributions and the practical application of the findings to the management of critical incidents and complaint handling within the hospitality sector.

References

Adamson, C., (1993) Evaluating complaint procedures, *Managing Service Quality*, vol. 3, no. 2, pp.439-48.

Armistead, C., Clark, G., Stanley, P. (1993) *Managing Service Recovery*, Cranfield School of Management, Cranfield.

Bailey, D., (1994), Recovery from customer service shortfalls, *Managing Service Quality*, vol. 4, no. 6, pp.25-8.

Baker, D. A. and Crompton, J. L., (2000) Quality, satisfaction and behavioural intentions. *Annals of Tourism Research*, vol. 27 no. 3, pp. 785 - 804.

Bell, C. R., & Zemke, R. E., (1992) *Managing knock your socks off service*, New York
Berry, L. L., (1995) *On great service: A framework for action*. New York: The Free Press.
Bitner, M., (1990) Evaluating service encounters: The effect of physical surroundings and employee responses, *Journal of Marketing,* vol. 4, no. 2, pp. 69 – 82.
Bitner, M.J., Booms, B.H., and Tetreault, M.S., (1990) The service encounter: diagnosing favourable and unfavourable incidents, *Journal of Marketing*, vol. 54, pp. 71-84.
Bitner, M.J., Booms, B.H., and Mohr, L.A., (1994) Critical service encounters: the employees' viewpoint, *Journal of Marketing*, vol. 58, pp. 95-106.
Blodgett, J. G. and Granbois, D. H., (1992) Toward an integrated conceptual model of consumer complaining behavior, *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, vol. 5, pp. 93-103.
Blodgett, J., Granbois, D., and Walters, R., (1993) The effects of perceived justice on complainants' negative word-of-mouth behaviour and re-patronage intentions, *Journal of Retailing*, vol. 69, no.4, pp. 399-427.
Blodgett, J. G., Wakefield, K. L., and Barnes, J. H., (1995) The effects of customer service on consumer complaining behaviour. *Journal of Services Marketing*, vol. 9, no. 4, pp. 31-42.
Bolfing, C.P., (1989) How do customers express dissatisfaction and what can service marketers do about it? *Journal of Services Marketing*, vol. 3, pp.5-23.
Boshoff, C., (1997) An experimental study of service recovery options, *International Journal of Service Industry Management*, vol. 8, no. 2, pp. 110 – 130.
Boshoff, C., and Leong, J., (1998) Empowerment, attribution and apologising as dimensions of service recovery An experimental study, *International Journal of Service Industry Management,* vol. 9, No. 1, pp. 24 – 47.
Bossone, K., (1994) Service passwords. *Executive Excellence*, Vol. 11, No. 9, pp. 18.
Campbell, L., and Finch, E., (2004). Customer satisfaction and organisational justice, *Facilities*, vol. 22, nos. 7/8, pp. 178 – 189.
Carman, J., (1990) Consumer perceptions of service quality: An assessment of the SERVQUAL dimensions. *Journal of Retailing*, Spring, no. 66, pp. 32 - 48.
Churchill, G. A., (1979) A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research,* 16 (February), pp. 64-73.
Churchill, G. A. and Surprenant, C., (1982) An investigation into the determinants of customer satisfaction. *Journal of Marketing Research*, vol. 19, pp 491 - 504.
Clark, G.L., Kaminski, P.F., and Rink, D.R., (1992) Consumer complaints: advice on how companies respond based on an empirical study, *Journal of Services Marketing*, vol. 6, no.1, pp.41-50.
Conlon, D. E., and Murray, N. M., (1996) Customer perceptions of corporate responses to product complaints: The role of explanations. *Academy of Management Journal*, vol. 39, no. 4, pp. 1040-1056.
Crompton, J. L., (1999) Service quality: the journey so far. Keynote Presentation, *ATLAS Annual Conference: 'Service Quality and Customer Service in Tourism and Leisure'*, Munich, 9-11 Sept.
Crompton, J.L. and Love, L. L., (1995) The predictive value of alternative approaches to evaluating quality of a festival. *Journal of Travel Research*, vol. 34, no. 1, pp. 11 - 24.
Cronin, J., and Taylor, S., (1992) Measuring service quality: a re-examination and extension, *Journal of Marketing,* vol. 56, no. 3, pp. 55 - 68.

Davidow, M., (2000) The bottom line impact of organisational responses to customer complaints, *Journal of Hospitality and Tourism Research,* vol. 24, no. 4, pp. 473 – 490.

Day, R. L. and Ash, S. B., (1979) Consumer response to dissatisfaction with durable products, in W. Wilkie (ed.), *Advances in Consumer Research*, vol. 6, Ann Arbor, MI: Association for Consumer Research, pp. 438-444.

Day, R. L. and Landon, Jr., E. L., (1977) Toward a theory of consumer complaining behaviour. In A. Woodside, J. Sheth and P. Bennett (eds.), *Consumer and Industrial Buying Behavior*, Amsterdam: North Holland Publishing Company, pp. 425-437.

Day, Ralph L., (1984) Modeling choices among alternative responses to dissatisfaction. In Kinnear, T.C. (ed.) *Advances in Consumer Research* 11. Provo, UT: Association for Consumer Research, pp. 496-499.

Dorfman, P. W., (1979) Measurement and meaning of recreation satisfaction: a case study in camping. *Environment and Behaviour*, vol. 11, no. 4, pp. 483 - 510.

Duke, C.R. and Persia, M.A., (1996) Consumer-defined dimensions for the escorted tour industry segment: expectations, satisfactions and importance. *Journal of Travel and Tourism Marketing*, vol. 5, nos. 1/2, pp. 77 - 99.

Eagly, A.H. and Chaiken, S., (1993) *The Psychology of Attitudes*. Orlando, FL: Harcourt Brace Jovanovich.

Ekinci, Y., Riley, M. and Chen, J. S., (2000) A review of comparison standards used in service quality and customer satisfaction studies: some emerging issues for hospitality and tourism research. In J.A. Mazanec, G.I. Crouch, J.R. Brent Ritchie and A.G. Woodside (eds.) *Consumer Psychology of Tourism, Hospitality and Leisure, Vol. 2*., Oxford, CABI, pp. 321-332.

Fallon, P. and Schofield, P., (2004) A comparison of models used in the measurement of tourist satisfaction. In Williams, J.A. and Uysal, M. (eds.) *Current Issues and Development in Hospitality and Tourism Satisfaction*, New York: Haworth Press, pp. 77-96.

Fick, G. R. and Ritchie, J. R. B., (1991) Measuring service quality in the travel and tourism industry. *Journal of Travel Research*, vol. 29 (Fall), pp. 2 - 9.

Firnstahl, T.W., (1989) My employees are my service guarantee, *Harvard Business Review*, vol. 67, Nno.4, pp.28-34.

Folger, R. & Cropanzano, R., (1998). *Organizational justice and human resource management*. Thousand Oaks: Sage.

Folger, R., and Greenberg, J., (1985) Procedural justice: an interpretative analysis of personnel systems. In Rowland, K.M., Ferris, G.R. (eds.), *Research in Personnel and Human Resource Management*, JAI Press, Greenwich, CT, vol. 3, pp.141-83.

Fornell, C. and Wernerfelt, B., (1988) A model of customer complaint management, *Marketing Science*, vol. 7, no. 3, (Summer), pp. 287-298.

Fornell, C. and Westbrook, R., (1979) An exploratory study of assertiveness, aggressiveness, and consumer complaining behaviour. In W. Wilkie (ed.), *Advances in Consumer Research,* Vol. 6, Ann Arbor, MA: Association for Consumer Research, pp. 105-110.

Garrett, D. E., Meyers, R. A., and Camey, J., (1991) Interactive complaint communication: A theoretical framework and research agenda. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behaviour*, vol. 4, pp. 62-79.

Getty, M.J. and Thomson, N.K., (1994) The relationship between quality, satisfaction, recommending behaviour in lodging decisions. *Journal of Hospitality and Leisure Marketing*, vol. 2, no. 3, pp. 8.

Geva, A. and Goldman, A., (1991) Satisfaction measurement in guided tours. *Annals of Tourism Research*, vol. 18, pp. 177 - 185.
Goodman, J. A., Malech, A., and Boyd, S., (1987) Danger, angry customer. *ABA Banking Journal*, vol. 79, no. 1, pp. 63-66.
Goodwin, C., and Ross, I., (1989) Salient dimensions of perceived fairness in resolution of service complaints. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behaviour*, vol. 2, pp. 87-92.
Godwin, B., Patterson, P. G. and Johnson, L. W., (1995) Emotion, coping and complaining propensity following a dissatisfactory service encounter, *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior,* vol. 8, pp. 155-163.
Granbois, D., Summers, J. and Frazier, G., (1977) Correlates of consumer expectations and complaining behavior. In. R. L. Day (Ed.) *Consumer Satisfaction, Dissatisfaction and Complaining Behaviour.* Bloomington, IN: Indiana University School of Business, pp. 18-25.
Hart, C.W.L., (1988) The power of unconditional guarantees, *Harvard Business Review*, vol. 66, no.4, pp.54-62.
Hart, C., Heskett, J., and Sasser, W., (1990) The profitable art of service recovery, *Harvard Business Review,* vol. 68, July – August, pp. 148 – 156.
Hewstone, M., (1989) *Causal Attribution: From Cognitive Processes to Collective Beliefs*. Oxford: Basil Blackwell.
Hoffman, K., Kelley, S., and Rotalsky, H., (1995) Tracking service failures and employee recovery efforts, *Journal of Services Marketing,* vol. 9, no. 2, pp. 49 – 61.
Hoffman, K., and Kelley, S., (2000) Perceived justice needs and recovery evaluation: A contingency approach, *European Journal of Marketing,* vol. 34, no. 3/4, pp. 418 – 433.
Huan, T., Beaman, J. and Shelby, L. B., (2002) Using action-grids in tourism management. *Tourism Management*, vol. 23, pp. 255-264.
Hughes, K., (1991). Tourist satisfaction: a guided tour in North Queensland. *Australian Psychologist*, vol. 26, no. 3, pp. 168- 178.
Jenks, M., (1993) Ten easy steps to defusing anger, *Mobius*, *12*(2), 36-39. Joreskog, K. G.,& Sorbom, D. (1996) *LISREL 8: User's reference guide*. Chicago: Scientific Software International.
Johnston, R., (1994) *Service Recovery: An Empirical Study*, Warwick: Warwick Business School.
Karatepe, O., and Ekiz, E., (2004) The effects of organisational responses to complaints on satisfaction and loyalty: A study of hotel guests in Northern Cyprus, *Managing Service Quality,* vol. 14, no. 6, pp. 476 – 486.
Keaveney, S.M., (1995) Customer switching behaviour in service industries: an exploratory study, *Journal of Marketing*, vol. 59, no.2, pp.71-89.
Kelley, S.W, Hoffman, and K.D, Davis, M.A., (1993) A typology of retail failures and recoveries, *Journal of Retailing*, vol. 69, pp. 429-52.
Khan, P., (2004) Measuring what matters: 'Importance' in Satisfaction Surveys. *Insights Tourism Intelligence Papers*.
Kozak, M., (2001) A review of comparison standards used in service quality and customer satisfaction studies: some emerging issues for hospitality and tourism research, in J.A. Mazanec, G.I. Crouch, J.R.B. Ritchie and A.G. Woodside (eds.) (2001) *Consumer Psychology of Travel, Hospitality and Leisure,* Vol. 2. Oxon: CABI, pp 321-332.

Kozak, M. and Nield, K., (1998) Importance-performance analysis and cultural perspectives in Romanian Black Sea resorts. *Anatolia: an International Journal of Tourism and Hospitality Research,* vol. 9, no. 2, pp. 99 - 116.
Laws, E., (1995). *Tourism Destination Management - Issues, Analysis and Policies*. London: Routledge.
Leblanc, G., (1992) Factors affecting customer evaluation of service quality in travel agencies: an investigation of customer perceptions, *Journal of Travel Research*, vol. 30 (Spring), pp. 10-16.
Lewis, B., and Clacher, E., (2001) Service failure and recovery in UK theme parks: The employee's perspective, (2001) *International Journal of Contemporary Hospitality Management,* vol. 13, no. 4, pp. 166 – 175.
Lewis, B., and McCann, P., (2004) Service failure and recovery: Evidence from the hotel industry, *International Journal of Contemporary Hospitality Management,* vol. 16, no. 1, pp. 6 – 17.
Lewis, B., and Spyrakopoulos, S., (2001) Service failures and recovery in retail banking: The customer's perspective, *International Journal of Bank Marketing,* vol. 19, no. 1, pp. 37 – 48.
Lewis, R. C., (1983) Consumers complain - What happens when business responds? In R. L. Day & H. K. Hunt (Eds.), *International fare in consumer satisfaction and complaining* (pp. 88-94). Bloomington, IN: Bureau of Business Research.
Litvin, S. W. and Ling, S., (2001) The destination attribute management model: an empirical application to Bintan, Indonesia. *Tourism Management*, vol. 22, pp. 481-492.
Mannell, R. C. and Iso-Ahola, S.E., (1987) Psychological nature of leisure and tourism experience. *Annals of Tourism Research,* vol. 14, no. 2, pp. 314 - 331.
Martilla, J. A. and James, J. C., (1977) Importance-performance analysis. *Journal of Marketing*, vol. 41, no. 1, pp. 77-79.
Martin, C. L., and Smart, D. T., (1994) Consumer experiences calling toll-free corporate hotlines. *Journal of Business Communications*, vol. 31, no. 3, pp. 195-212.
Mason, C.J., (1993) What image do you project, *Management Review*, no.82, pp.10-16.
Mattila, A., (2004) The impact of service failures on customer loyalty, *International Journal of Service Industry Management,* vol. 15, no. 2, pp. 134 – 149.
Maxham, J., (2001) Service recovery's influence on consumer satisfaction, positive word of mouth and purchase intentions, *Journal of Business Research,*v. 54, no. 1, pp. 11 – 24.
McDougall, G., and Levesque, T., (1999) Waiting for service: The effectiveness of recovery strategies, *International Journal of Contemporary Hospitality Management,* vol. 11, no. 1, pp. 6 – 15.
Meyer, A. and Westerbarkey, P., (1996) Measuring and managing hotel guest satisfaction. In D.M. Olsen, R. Teare and E. Gummesson (eds.) *Service Quality in Hospitality Organisations*. New Ork: Cassel. pp. 185-204.
Meyer M. W., (1997) Dilemmas of performance management, *Journal of Strategic Performance Management*, (April-May) pp.33-42.
Mill, R. C., (1989) Customizing Government Research: the Use of Perceptual Maps. *Proceedings of the Travel and Tourism Research Association*, Twentieth Anniversary Conference. Salt Lake City, UT: Bureau of Economic and Business Research, University of Utah, pp. 109 - 114.
Miller, J. A., (1977) Studying satisfaction: Modifying models, eliciting expectations, posing problems, and making meaningful measurements. In *Conceptualisation and*

Measurement of Consumer Satisfaction and Dissatisfaction. Bloomington: Indiana School of Business. pp. 72 - 91.
Miller, J.L., Craighead, C.W., and Karwan, K.R. (2000) Service recovery: a framework and empirical investigation, *Journal of Operations Management*, no.18, pp.387-400.
Morris, S. V., (1988) How many lost customers have you won back today? An aggressive approach to complaint handling in the hotel industry. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behaviour*, vol. 1, pp. 86-92.
Moyer, M. S., (1984) Characteristics of consumer complainants: implications for marketing and public policy, *Journal of Public Policy and Marketing*, vol. 3, pp. 67-84.
Nyquist, J.D., Bitner, M.J., and Booms, B.H., (1985) Identifying communication difficulties in the service encounter: a critical incident approach, in Czepiel, J., Solomon, M., Surprenant, C. (eds), *The Service Encounter*, Lexington, MA: Lexington Books. pp.195-213.
Oh, H. (1999) Service quality, customer satisfaction, and customer value: an holistic perspective. *Hospitality Management*, vol. 18, pp. 67 - 82.
Oh, H. and Parks, C. S., (1998) Evaluating the role of attribute importance as a multiplicative weighting variable in the study of hospitality consumer decision-making. *Journal of Hospitality & Tourism Research*, vol. 21, no. 3, pp. 61-80.
Oliver, R.L., (1997) *Satisfaction: A Behavioral Perspective on the Consumer*. New York: McGraw-Hill International Editions.
Orams, M.B. and Page, S.L., (2000) Designing self-reply questionnaires to survey tourists: issues and guidelines for researchers. *Anatolia*, vol. 11, no. 2, pp. 125-139.
Ostrom, A., and Iacobussi, D., (1998) The effect of guarantees on consumers' evaluation of services, *Journal of Services Marketing,* vol. 12, no. 5, pp. 362 – 378.
Pearce, P.L., (1991) Introduction to the Tourism Psychology. *Australian Psychologist*, vol. 26, no. 3, pp. 145 -146.
Pizam, A. and Ellis, T., (1999) Customer satisfaction and its measurement in hospitality enterprises. *International Journal of Contemporary Hospitality Management,* vol. 11, no. 7, pp. 326 -339.
Plymire, J., (1991). Complaints as opportunities. *The Journal of Services Marketing*, vol. 5, no. 1, pp. 61-65.
Reichheld, F.F., and Sasser, W.E., (1990) Zero defections: quality comes to services, *Harvard Business Review*, vol. 68, no.5, pp.105-11.
Sampson, S. E. and Showalter, M. J., (1999) The performance-importance response function: observations and implications. *The Service Industries Journal*, vol. 19, no. 3, pp. 1-25.
Schofield, P., (2001) Evaluating heritage visitor attractions from the consumer perspective: a focus on Castlefield Urban Heritage Park in Manchester, U.K. In J.A. Mazanec, G.I. Crouch, J.R.B. Ritchie and A.G. Woodside (eds.) *Consumer Psychology of Travel, Hospitality and Leisure, Volume 2*. London: CAB International. pp. 285 – 302.
Schofield, P. and Fallon, P., (2000) Measuring the importance of critical factors in restaurant service quality performance evaluation: a triadic perspective. *Proceedings of the 1st Consumer Satisfaction Research in Tourism and Hospitality Conference*, Oxford Brookes University, Oxford, 25 Nov.
Singh, J., (1988) Consumer complaint intentions and behaviour: Definitional and taxonomical issues. *Journal of Marketing*, vol. 52 (January) pp. 93-107.

Singh, J., (1990) Voice, exit, and negative word-of-mouth behaviours: An investigation across three service categories, *Journal of the Academy of Marketing Science,* vol. 18, no. 1, pp. 1-15.
Smith, A. K., and Bolton, R. N. (1998) An experimental investigation of customer reactions to service failure and recovery encounters: Paradox or peril? *Journal of Service Research*, vol. 1, no. 1, pp. 65-81.
Smith, A.K., Bolton, R.N., and Wagner, J., (1999) A model of customer satisfaction with service encounters involving failure and recovery, *Journal of Marketing Research*, vol. 36 no.3, pp. 356-89.
Smith, C. A., and Ellsworth, P.C., (1985) Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology,* vol. 48, pp. 813 – 838.
Sparks, B., and Callan, V., (1995) Dealing with service breakdowns: the influence of explanations, offers and communication style on consumer complaint behaviour, *Proceedings of the World Congress 7th Bi-annual Marketing Science Conference,* vol. 7, nos. 2, 9, pp. 106 – 115.
Sparks, B A and McColl-Kennedy, J.R., (2001) Justice strategy options for increased customer satisfaction in a services recovery setting, *Journal of Business Research,* vol. 54, pp. 209-218.
Spreng, R. A., Harrell, G. D., and Mackoy, R. D., (1995) Service recovery: Impact on satisfaction and intentions. *Journal of Services Marketing*, vol. 9, no. 1, pp. 15-23.
Tax, S. S., and Chandrashekaran, M., (1992) Consumer decision making following a failed service encounter: A pilot study. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behaviour*, vol. 5, pp. 55-68.
Technical Assistance Research Programs., (1986) *Consumer Complaint Handling in America: An Updated Study*. Washington, DC: Office of Consumer Affairs, Technical Assistance Research Programs.
Testa, M. R., Skaruppa, C., and Pietrzak, D., (1998) Linking job satisfaction and customer satisfaction in the cruise industry: Implications for hospitality and travel organizations. *Journal of Hospitality & Tourism Research,* vol. 22, pp. 4-14.
Tyler, T., (1994). Psychological models of the justice motive: Antecedents of distributive and procedural justice, *Journal of Personality and Social Psychology,* Vol. 67, No. 5, pp. 850 – 863.
Walsh, A., (1996) Customer perception of quality of service and response to the complaint handling process of a rental car company, *Journal of Consumer Satisfaction, Dissatisfaction, and Complaining Behaviour*, vol. 9, pp. 190-199.
Weiner, Y., (1982) Commitment in the organisation: a normative view, *Academy of Management Review,* vol. 7, no. 3, pp. 418 – 428.
Westbrook, R. A. and Oliver, R. L., (1991) The dimensionality of consumption emotion patterns and consumer satisfaction, *Journal of Consumer Research*, vol. 18, pp. 84 – 91.
Weun, S., Beatty, S., and Jones, M., (2004) The impact of service failure severity on service recovery evaluations and post-recovery relationships, *Journal of Services Marketing,* vol. 8, no. 2, pp. 133 – 146.
Whitely, R. (1994) Fix the customer first. *Sales and Marketing Management*, vol. 146, no. 8, pp. 49-50.
Yuksel, A. and Rimmington, M., (1998) Customer satisfaction measurement, *Cornell Hotel and Restaurant Administration Quarterly*, Dec, pp 60 -70.
Zeithaml, V., Parasuraman, A. and Berry, L., (1985) *Delivering Service Quality*. New York: Free Press.
Zemke, R., (1994) Service recovery. *Executive Excellence*, vol. 11, no. 9, pp. 17.

Negative Refraction in Perspective

A D Boardman, N King and L Velasco
Institute for Materials Research

Introduction

There is a rapidly growing literature that uses the word *negative* (Pendry 2003, 2004; Lahthakia 2003a, 2003b; McCall, Lahthakia and Weiglhofer 2002; Lindell et al. 2001; Ziolkowski and Kipple 2003; Smith, Schurig and Pendry 2002; Boardman et al. 2005) in connection with phase velocity, or refraction. In addition, this adjective is quite often associated with the materials that are called *left-handed* and the refraction they display. Many articles appear to discuss the topic of this paper in the correct kind of way but some go on to misdirect by inappropriate associations embedded in what is often an inappropriate terminology. On the face of it, it is a daunting task to give a satisfactory perspective on the topic of negative refraction so perhaps it is wise to take a leaf out of Lewis Carroll's book "Alice's Adventures in Wonderland" and simply take the view that it is best to begin at the beginning, go on until the end is reached and then stop. The temptations to engage in elegant but abstract descriptions of a theoretical kind, or just encyclopaedic descriptions, will be avoided in favour of a selective offering that makes important points. The arguments are visually supported, sometimes at a very advanced computational level.

The beginning belongs to the great optical scientist Sir Arthur Schuster who, as Professor of Physics in the University of Manchester, in 1904, was a colleague of the famous hydrodynamicist Sir Horace Lamb. It is not possible to be absolutely categorical about the conversations that they had but the outcome that matters to us is crystal clear. Lamb was then concerned about the relationship between the group velocity (Lamb 1916) and individual, or phase, velocity of the waves bundled together as a group. First of all, it is very easy to witness group and phase velocity in action, simply by throwing a pebble into a pond, or lake, and then watching the ripples spread out. The disturbance appears to be limited in space and time and is in fact a bunch, or group, of faster or slower waves that then travel through it. In practice, these individual waves appear at one end of the group, travel across it and then simply die out at the other end. It is intuitively obvious that energy is carried by the group and that this energy is being transported at the group velocity. Lamb also recognised that the individual waves in a group can, in principle, also set out at the front of the group and die out at the rear of it. Schuster then considered (Schuster 1904) this type of *backward wave* phenomenon in an optical context. He pointed out that backward waves may be generated whenever light propagates within an optical absorption band. In such regions, what came to be known, inappropriately, as anomalous dispersion occurs. All materials have one or more bands like this, however, so there is nothing unusual about materials like fuchsin and cyanin (figure 1) that were nominated by Schuster as examples of this phenomenon. In an absorption region the wave velocity increases as the wavelength decreases and the group velocity can be anti-parallel to the phase (individual) velocity. However, Schuster did not discuss the kind of possibilities outlined below.

Schuster found this all very fascinating but was unimpressed by the fact that massive absorption implies that material thicknesses of less than a wavelength would be forced upon any experimental test for backward waves. Apart from this caveat, Schuster produced an impressive diagram (Figure 2) showing what will happen when a plane wave of light is incident upon an interface of a semi-infinite backward wave medium. Schuster states that "One curious result follows: the deviation of the wave entering such a medium is greater than the angle of incidence".

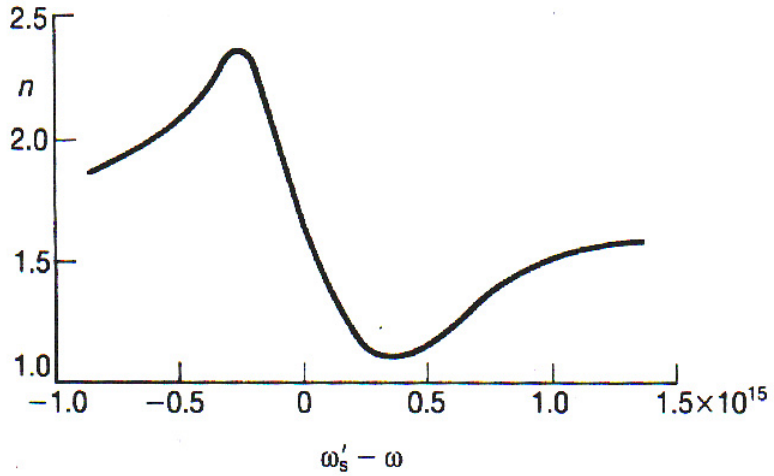


Figure 1: Dispersion curve for cyanin showing a region of anomalous dispersion. Note that the curve is displayed by taking the origin at a frequency $\omega_s^{'}$

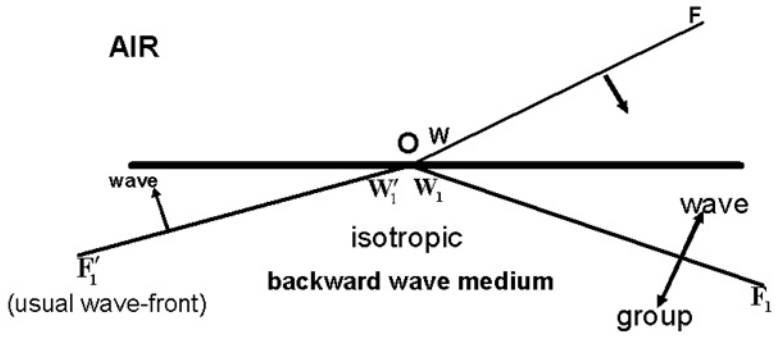


Figure 2: A redrawn, edited, form of Schuster's original diagram in which he shows the incident wave front WF. He also shows the direction of propagation, in the usual way, by the arrow that is perpendicular to a wave front.

There are two possibilities for the refracted wave front i.e. W_1F_1 or $W_1^{'}F_1^{'}$. Schuster noted that the problem with using $W_1^{'}F_1^{'}$ is that its intersection with the surface slides to the left while the intersection of the incident wave front slides to the right. This is because the material sustains phase waves that are propagating towards the surface. In order to lock these intersection points together it is necessary to adopt the refraction wave front W_1F_1. In this way, Schuster was able to conclude that energy can be carried forward at the group velocity but in a direction that is anti-parallel to the phase velocity. Given this conclusion, an incident light wave looks as though it refracts past

the normal and this aroused Schuster's curiosity. It is interesting that this conclusion was reached for a purely dielectric material, characterised by relative dielectric permittivity and a relative permeability equal to that of free space. In other words, he saw that there was no necessity to invoke a dispersive relative magnetic permeability, as is the case for the so-called left-handed media to be addressed below.

Energy is transported at the speed of the group and the convention that it is being transported in a "*positive*" direction can be taken. It is then logical to assert that in the *isotropic* medium discussed by Schuster that the phase velocity is travelling in a "*negative*" direction. Of course, there is a kind of upbeat arbitrariness to this convention because all that is being conveyed is that in an isotropic backward wave medium the energy flow and the phase velocity direction are anti-parallel. Nevertheless, the advocacy that this type of isotropic medium should be referred to as a *negative phase velocity* medium carries considerable conviction. A second point concerns the refraction itself. Again it is not unreasonable to refer to the normal, everyday, refraction that everybody has witnessed as a manifestation of *positive refraction* so that any refraction that goes beyond the normal could be called *negative refraction*. It does not follow that it is necessary to describe this behaviour in terms of an actual negative refractive index, but this issue will be returned to later on in the paper. In the Schuster example, negative refraction is clearly linked to the negative phase velocity, discussed earlier. However, much of the modern literature has taken the two things to be entirely synonymous. This is not always the case and it will be seen later that there are very common examples of negative refraction, in which the phase wave and group travel in a *forward*, but not necessarily the same, direction. A note of caution then is to avoid always associating negative refraction with the presence of backward waves.

Now that some of the background has been established this paper will go on to address a number of negative refraction examples and will highlight some of the controversy that created such a lot of publicity for this area of metamaterials. To complete the introduction it is necessary to move on nearly half a century from Schuster to 1945. This year sees Mandelshtam setting out the rules obeyed by negative refractive media (Mandelshtam 1945), although he is apparently unaware of the pioneering work of Schuster. One implication of negative refraction is the possibility of creating perfect lenses and much has been made of this in the last few years through some very elegant work (Pendry 2003). Basically, negative refraction permits the use of lenses with plane parallel sides (Silin 1978). In that case monochromatic aberrations can be eliminated. This dramatic consequence has received a lot of attention in recent times by Pendry. This type of lens behaviour is a direct consequence of negative refraction and will not be addressed further in this paper. The aim instead is to establish the basic nature of negative refraction and discuss any conceptual difficulties surrounding it, rather than elaborate particular applications.

Negative Phase aka Left-Handed Materials

Schuster pointed to an absorption band as a possible medium in which to generate backward waves but was concerned that it is a region of high absorption. He did not need to invoke a negative relative permittivity nor, indeed, the kind of relative permeability normally associated with purely magnetic materials. The creation of backward waves, however, also flows as an elementary consequence of Maxwell's

equations for a lossless, *isotropic* medium. What is required, first of all, is a dispersive relative permittivity $\varepsilon(\omega)$ coupled to a dispersive relative permeability $\mu(\omega)$, where w is the angular frequency. The next step is to set $\varepsilon(\omega)$ and $\mu(\omega)$ to be *simultaneously negative* (Veselago 1968) in a certain frequency range to create a transparent band of operation.

As discussed by Veselago, adopting *scalar functions* $\varepsilon(\omega)$ *and* $\mu(\omega)$, simultaneously setting them to be *negative*, and then proceeding with a plane wave solution to Maxwell's equations leads to some elementary conclusions. Assume that plane waves $\exp i(\mathbf{k.r} - \omega t)$ are propagating, where $\mathbf{k}$ is the wave vector, $\mathbf{r}(x, y, z)$ is the spatial vector involving the coordinates (x, y, z), and t is the time coordinate. Two of Maxwell's equations then yield

$$\begin{aligned} \mathbf{k} \times \mathbf{E}(w) &= w m_0 m(w) \mathbf{H}(w) \\ \mathbf{k} \times \mathbf{H}(w) &= -w e_0 e(w) \mathbf{E}(w) \end{aligned} \tag{1}$$

where $\mathbf{E}(w)$ and $\mathbf{H}(w)$ are complex and are the Fourier transforms of the field vectors. Hence the Poynting vector $\mathbf{S}$, giving a measure of energy flow, is

$$\mathbf{S} = \frac{1}{2}\left(\frac{1}{w m_0 m(w)}\right)\mathbf{k}|\mathbf{E}|^2 = \frac{1}{2}\left(\frac{1}{w m_0 e(w)}\right)\mathbf{k}|\mathbf{H}|^2 \tag{2}$$

The immediate danger here is that this is only a plane wave calculation and it can lead to some confusion, when Snell's law is applied. This point will be put on one side until worries about causality are introduced further on in this article. For the moment, it is apparent that, under the condition $\mu(\omega) < 0$ and $\varepsilon(\omega) < 0$, $\mathbf{k}$ is *anti-parallel* to $\mathbf{S}$, and that *backward waves* can be expected in an isotropic medium for which $\mu(\omega) < 0$ and $\varepsilon(\omega) < 0$. Nothing has been assumed about the refractive index, nor is this strictly necessary since it is what can be called a *derived concept* and does not appear in Maxwell's equations. It can be set equal to the positive root $n = \sqrt{m(w)e(w)}$, or the negative root. The choice does not affect the necessary appearance of backward waves but it is merely derived from the fact that any square root offers the options ± (Pokrovsky and Efros 2002). A familiarity with vector analysis leads to a question about the vector products in (3). For $\mu(\omega) > 0$ and $\varepsilon(\omega) > 0$, a *right-handed* rotation about $\mathbf{H}(w)$ takes $\mathbf{k}$ to $\mathbf{E}(w)$. For $\mu(\omega) < 0$ and $\varepsilon(\omega) < 0$, however, it is a *left-handed* rotation about $\mathbf{H}(w)$ that takes $\mathbf{k}$ to $\mathbf{E}(w)$. The handedness refers to the rotation of this vector set and *not* any lack of symmetry in the material. In order to possess handedness (Jaggard, Mickelson and Papas 1979) a material must not be capable of being changed by a symmetry operation into its mirror image. Sugar solution is a familiar example of something exhibiting chirality, but then so is a golf club. To say that the left-handed (Veselago 1968) description of backward wave phenomena is a misunderstanding is an understatement but the literature has put this spin upon it and it will be hard to displace. In some ways, it is a similar development to the use of anomalous dispersion, over the last century, when it is plainly not anomalous at all. A much better description is to say it is negative phase medium (Lakhtakia 2003b).

An illustration of the use of negative phase media in refractive situations will now be shown. Each figure is the outcome of a finite-difference time-domain (FDTD) simulation (Taflove and Hagness 2000). First, the electromagnetic equivalent of dropping a stone into a pool is shown in Fig.3. A point source is generated in positive

phase free space. The pulse spreads out as a forward wave so the phase waves are travelling in the same direction as the electromagnetic ripples. The pulse eventually interacts with an impedance-matched slab (Ziolkoswki 2003) of negative phase material modelled in the FDTD through a Drude model for the permittivity and permeability with losses set to zero. The spatial grid used is square, the Courant number equal to 0.5 and the source is a hard source located at (126,250). The linear dispersive left-handed, or negative phase, slab is located between x = 250 and 500 and there is no restriction along the y axis. Figure 3 shows that the effect of the negative phase slab is to make the already diffracting beam converge to a single point. This is because of negative refraction.

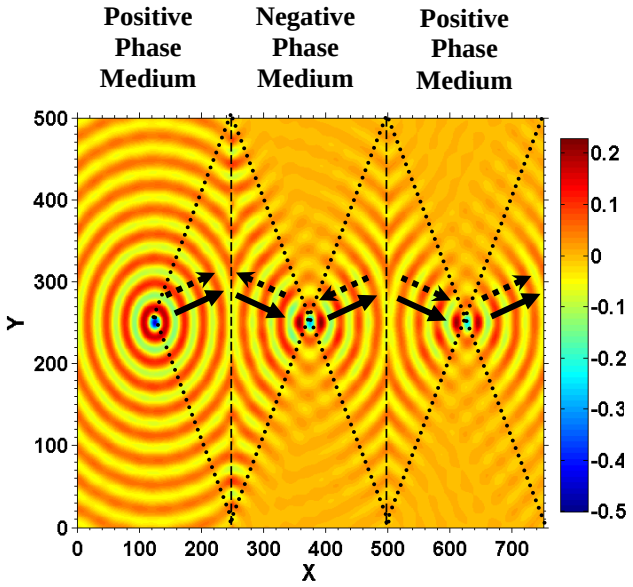


Figure 3: Point source located in positive phase medium interacting with a slab of negative phase material located between x = 250 and 500 for the whole y-axis, the source is located at the point (126,250). Group and phase velocities indicated by solid or dotted arrows respectively.

Another interesting example concerns the creation of surface plasmon polaritons (Boardman 1982), a special type of field trapped at an interface. Figure 4 shows this type of surface wave being launched using a thin film configuration (Kretschmann 1971) for an incident angle of 60º. The permittivity and permeability of the upper dielectric material are (e = 2.3409 m = 1), For the negative phase film e = -5.8349 m = -0.2900. The interfaces of this system are at y = 200 and 220. There is negative refraction involving backward waves.

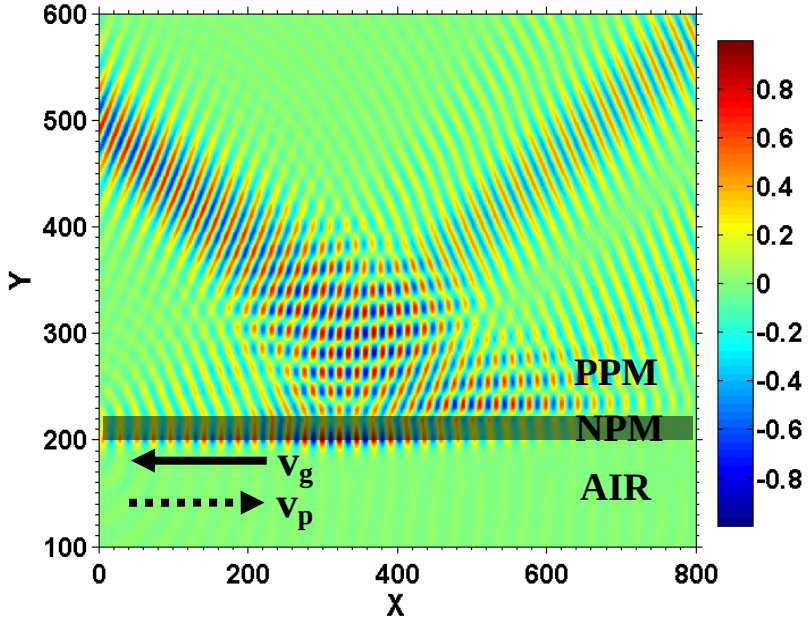


Figure 4: Surface wave formation using the Kretschmann configuration for an incident angle of 60°. PPM (e = 2.3409 m = 1), NPM (e = -5.8349 m = -0.2900). Interfaces at y = 200 and 220. Phase velocity is v_p and group velocity is v_g.

It must be expected that the underlying condition $\mu(\omega) < 0$ and $\varepsilon(\omega) < 0$ may have to be altered to take into account damping (McCall, Lahthakia and Weiglhofer 2002) that is modelled through the use of, respective, complex relative permeability and permittivity $\mu = \mu' + i\mu''$, $\varepsilon = \varepsilon' + i\varepsilon''$. One form for this condition found direct from $\varepsilon(\omega)$ and $\mu(\omega)$ is found to be

$$\left(\varepsilon' - |\varepsilon|\right)\left(\mu' - |\mu|\right) > \varepsilon''\mu'' \tag{3}$$

Or equivalently, since the imaginary part of k must remain positive so that backward waves occur (Weiglhofer and Lakhtakia 2003) a similar condition can also be found

$$\varepsilon'\mu'' + \varepsilon''\mu' < 0 \tag{4}$$

These criteria are displayed numerically in Figures 5 and 6 using the models (Ruppin 2000; Pendry et al. 1998; Smith et al. 2000)

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\Gamma\omega} \quad \mu(\omega) = 1 - \frac{F\omega^2}{\omega^2 - \omega_0^2 + i\Gamma\omega} \tag{5}$$

and data F = 0.56, $\omega_0 = 0.4\omega_p$, $\Gamma = 0.1\omega_p$.

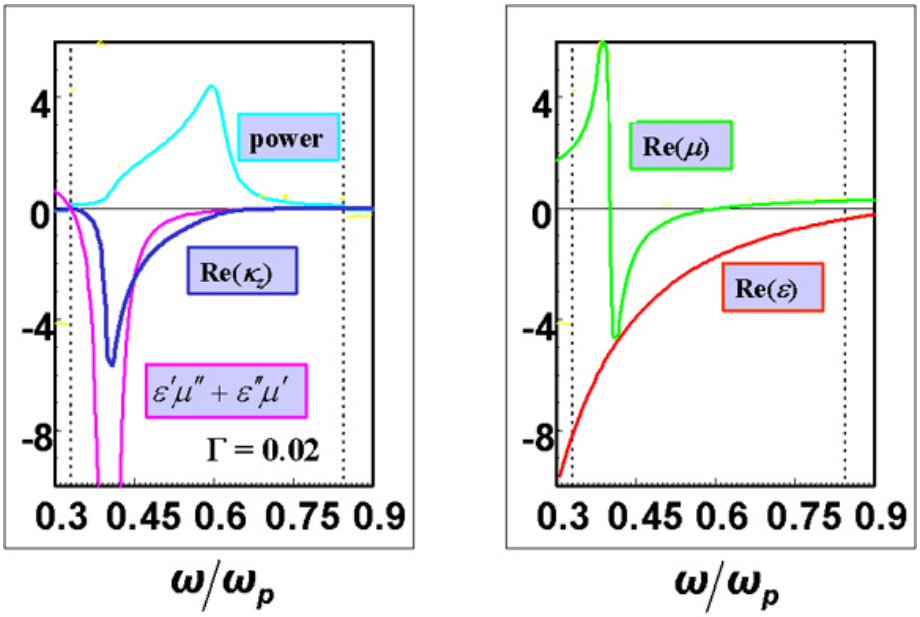


Figure 5: Γ=0.02 (a) Plots of the power flow, real part of the wave vector and the condition to have negative refraction (b) Real parts of e and m. Whenever e < 0 *and* m < 0 backward waves an occur

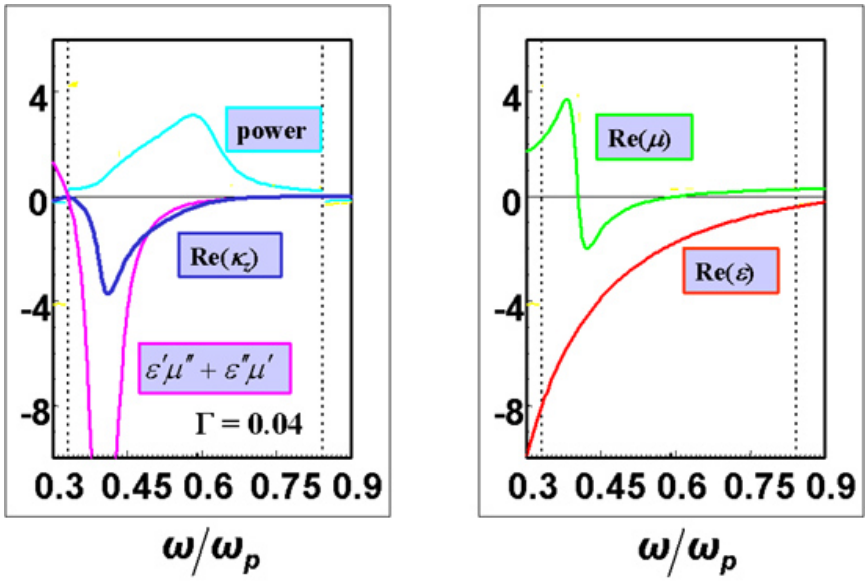


Figure 6: Γ=0.04 (a) Plots of the power flow, real part of the wave vector and the condition to have negative refraction (b) Real parts of e and m. When m> 0 and e <0 it is still possible to have backward waves

Causality

In the preceding discussions it has been tacitly accepted that the negative refraction is perfectly possible in principle and, indeed, observable if only the right kind of experiment can be done, or the right kind of material can be created. That negative refraction should be possible in the absorption bands of a whole host of materials was made clear by Schuster but the expectation of very high losses dampens any enthusiasm for this kind of experiment. If materials with negative relative permittivity and negative relative permeability could exist, or be made as a kind of exotic metamaterial, then backward waves and negative refraction should also be observed. The problem until recently has been to find such materials but the beautiful

experiments of Smith and co-workers and the elegant and persuasive papers by Pendry on the possibility of actually creating perfect lenses have provided physical evidence that such materials can be constructed. Yet, not too long ago, a dark cloud of doubt was thrown over the whole area when the very idea of negative refraction, and also the experimental evidence (Smith and Kroll 2000) was questioned (Valanju et al 2002) both on the grounds of interpretation and accuracy. This dark cloud concerns *causality* (Toll 1956) so this will now be examined in enough detail to reflect the various discussions that have appeared in the literature.

In macroscopic physics the principle that everything has a cause can be safely embraced. Developing the argument further involves the velocity of light and whether it is possible that you can travel backwards in time, or not. Travelling backwards in time has tricky consequences and science fiction writers are fond of using the grandfather paradox (Barjavel 1943), which goes like this. Suppose you travel back in time and prevent your biological grandfather from meeting your grandmother. Then you would have never been conceived, and just as in the film "Back to the Future" you would then have the problem that you would start to fade away through the *lack of a cause*. This paradox can be used to argue that travelling backwards in time must be impossible, if living entities and energy transportation is involved. The fact that no output can be expected before there is an input has been referred to as strict causality (Toll 1956) and can be easily appreciated by imagining that a source of electromagnetic wave is created. Switching on a source, like an aerial, and them pumping the waves out into free space will *cause* electromagnetic waves to propagate out towards an observer, some distance away. The *effect* upon this observer will be felt at some time later. It is not possible for the observer to access earlier times and, furthermore, the waves cannot travel faster than the velocity of light. Hence, if the observer experiences the waves at any earlier time then the golden rule of Einstein that energy cannot travel faster than the velocity of light will have been violated. It would also imply travelling backwards in time.

When the idea of negative refraction was discussed in modern times it sparked off quite a debate and it all centred upon the cherished principle of causality (Valanju et al. 2002). The relationship of causality to dispersion is well known and has been thoroughly discussed over the many decades but it is the connection to negative refraction that is so interesting; so this is the focus of attention here. Schuster, who first pointed to the possibility that light incident upon a block of a certain kind of material can be bent by refraction past the normal, was pre-Einstein when his book was published and made no comment, other than to say that it was a curious turn of events. The general principle that there can be materials that can support backward waves is acceptable, but what can be reasonably expected about the refraction process? Also, are the modern objections to negative refraction, apparently aimed at negative phase media, sustainable in any way?

The idea that refraction past the normal is a rational idea can be appreciated by revisiting the whole question but this time using the usual *phase-matching* graphical construction. A flat surface, of infinite extent, separating two semi-infinite isotropic half-spaces will now be considered and in each region the dispersion equation is simply

$$k^2 = \frac{\omega^2}{c^2}\varepsilon\mu \tag{6}$$

where k is the wave number, w is the angular frequency, e is the relative dielectric permittivity, m is the relative magnetic permeability and c is the velocity of light in a vacuum. Equation (13) defines a surface in *k*-space (Saleh and Teich 1991) and Fig. 7 shows that for the isotropic case being considered that the cross-sections are circles.

The wave vectors that are shown represent the *phase* directions. The normals to the surfaces, i.e. the arrows that are perpendicular to the tangents, show the expected direction of the group velocity and hence the energy flow direction. Diagrams like Fig 7 are often used to illustrate phase matching and for a negative phase medium it is the backward flowing phase that must be used to achieve the necessary matching. No phase matching can be achieved by using the forward flowing phase. Once this construction is accepted then the corollary is that the group velocity is anti-parallel to the phase direction and negative refraction is clearly demonstrated. This would appear to the end of the discussion at this point until it realised that this a phase matching diagram, based upon a plane wave and no proper discussion of energy flow has yet taken place. Intuitively, however, it is expected that negative refraction does take place and that the experiments (Smith et al 2000) are beyond reproach.

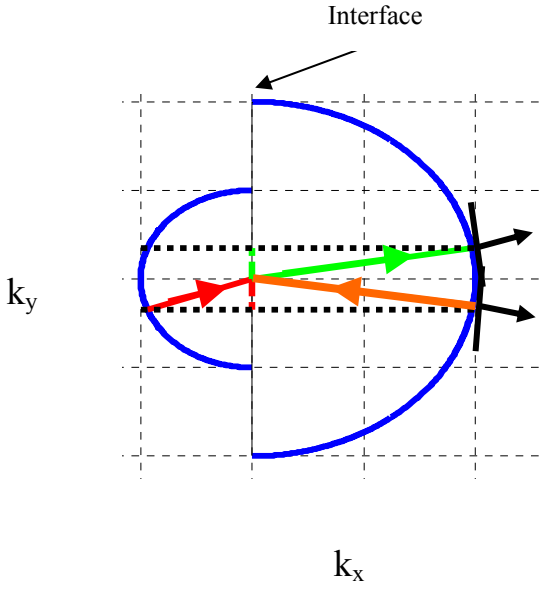


Figure 7: Wave number surfaces yield circles in the (k_x , k_y) space selected for this illustration. The interface separates air on the left from a negative phase medium on the right which is sustaining a negative phase velocity. The phase directions are shown and the normals to the tangents surfaces show the energy flow direction.

Explicitly, the objections raised a few years ago, to negative refraction, can be set out as follows. Figure 8 adopts the usual method of sketching how refraction works. On the left-hand side two *rays* [carrying energy] defining a beam of light incident upon a flat, unbounded surface are shown. Suppose that the beam approaches the surface through a medium for which $\mu(\omega)$ =1 and $\varepsilon(\omega)$ =constant > 0 and that it then refracts out into air.

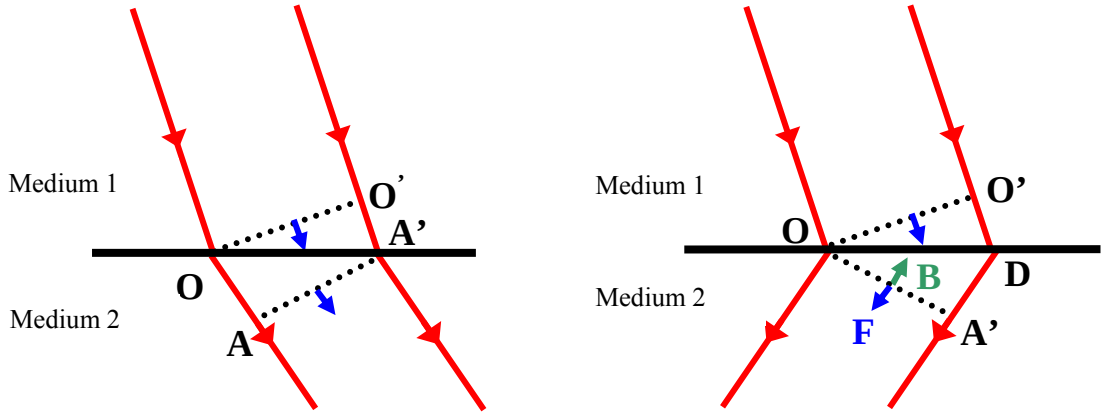


Figure 8: (a) the positive refraction expected when a beam of light encounters an interface between two positive phase isotropic dielectrics that support forward waves. The wave fronts are dotted lines. The arrows perpendicular to the fronts show the energy flow direction. (b) the same beam of light entering into a negative phase medium. Energy flow in the negative phase medium is in the direction labelled F and the label B denotes a backward phase wave.

The wave fronts are perpendicular to the ray directions and point O by the ray on the left. The ray on the right has to make up the distance O'A' but while it is doing this the first ray enters the medium and travels the distance OA. Thus, just at the moment when A' is in position, A is also lined up with it to form the front that is needed for positive refraction. This is entirely expected and there is no reason to suppose that it is unsafe to say that the energy flow is perpendicular to this wave front. The second case on the right is more problematical because the lower half-space consists of a negative phase medium for which $\mu(\omega) < 0$ and $\varepsilon(\omega) < 0$. For the new phase front to be developed in this negative phase medium the outer ray must travel through the distances O'D and DA' in zero time, in order to line up with O, to create the new phase front OA'. To put it bluntly, the point O' must get to A' with infinite speed and this violates the Einstein golden rule that the speed of light cannot be exceeded by an actual energy flow. This apparent violation that the speed of light is finite must also for the reasons given earlier violate causality too i.e. when A' matches up with O to form a front then what is a happening at A' is a cause without an effect. It is like travelling back in time. This is allowed for phase fronts but not for energy rays. The sketch in Figure 8 has been used to claim (Valanju 2002) that negative refraction is not permitted because of the causality violation and the answer to this challenge is quite interesting but it can be countered rather successfully.

On the simplest of levels, the elementary discussion based around a monochromatic plane wave is correct for the phase behaviour because *no energy is carried* by the phase velocity. It is simply enough to say that the negative phase medium carries backward waves that are *moving towards* the surface. This means that Figure 8 may be safely used for phase-matching arguments and the application of Snell's law. The energy flow is a different matter, because Snell's law does not apply to this distribution i.e. there is no phase diagram for the energy flow (Saleh and Teich 1991). Nevertheless, if a separate consideration of the Poynting vector is performed, it ought to be enough to say that the Poynting vector is anti-parallel to the phase vector and that it flows in the direction of the normal to the **k**-space frequency surfaces. To introduce group velocity directly into the description, however, it is necessary to consider more than one plane wave: a beam in other words. A linear beam being used

to interrogate a surface is, in fact, an infinite set of plane waves but even a limited set of plane waves may well be the basis of a interesting investigation to see how the energy flows are set up (Pendry and Smith 2002). These authors countered the objection raised to negative refraction by considering the interference pattern created by adding two plane waves, with slightly different frequencies, that are travelling in slightly different directions. This produces a fascinating effect in which the "group" of two waves upon encountering a negative phase medium transmit their energy with a group velocity that is anti-parallel to the phase wave direction but there is also an interference front. To explore this further, figure 9 shows the result of rigorous simulation that embodies both time and space modulation.

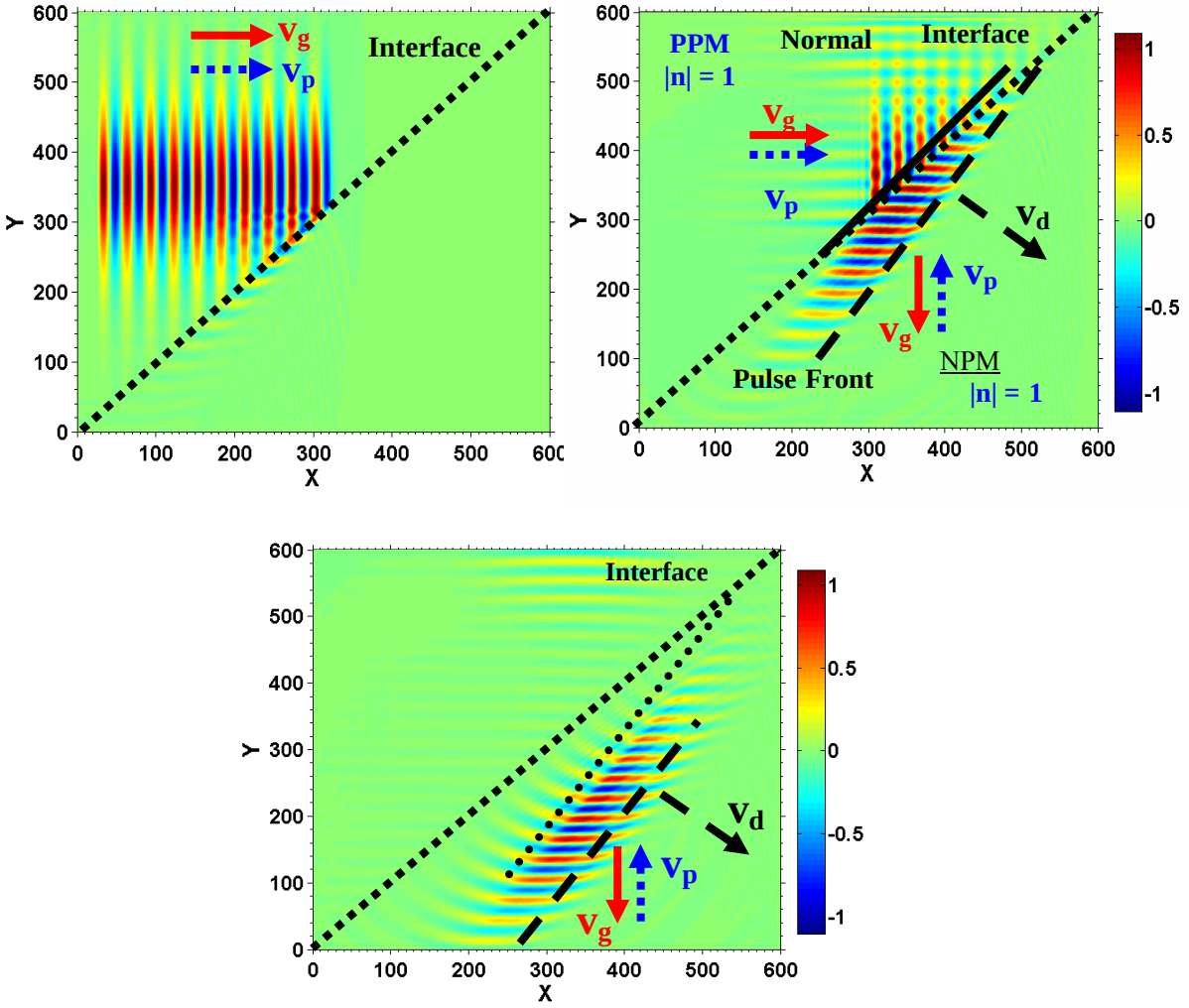


Figure 9: FDTD simulation of a space- time-modulated excitation as it crosses an interface between air and a negative phase medium. For convenience the interface is impedance matched. The figures are snapshots of how the input pulse is evolving for three different times. Note that this type of pulse possesses a large number of frequency components.

It is perhaps worth re-emphasising that these FDTD results are powerful and rely only upon unadulterated Maxwell's equations and record time and space outcomes in a dramatic fashion. Accordingly, Figure 9 reveals (Pendry and Smith 2002; Maslovski 2002) that an incoming pulse changes its shape as it goes through an interface. For

greater clarity, Figure 9 opts for an interface between a positive and negative phase medium that is inclined at an angle to the incoming space-time pulse. It is important when dealing with beams to distinguish between time-modulation that produces a front that moves in a positive refraction direction and space-modulation that produces negative refraction of the actual energy flow. Figure 9 shows precisely this turn of events. The pulse does become distorted. The energy flow, labelled with an arrow showing the direction of the group velocity, is in the negative refraction direction. Finally, the front associated with the time-modulation propagates with a velocity v_d that is set at an angle to the group velocity vector.

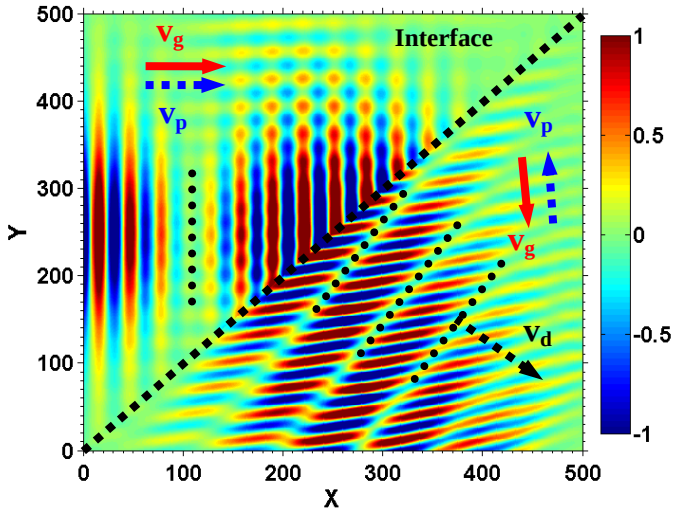


Figure 10: Incident modulated beam in which there are two predominant carrier frequencies. The interface is between air and a negative phase medium. Note the highly visible subsequent *positive* refraction of the temporal interference fringes. The minima have been extenuated by dotted lines.

In Figures 9 and 10 it can be seen that the FDTD method also produces the phase waves and that these are flowing anti-parallel to the group velocity. This is the classic hallmark of backward wave behaviour. All of this discussion points to the fact that the original objection to negative refraction in negative phase media cannot be sustained. The outcome shown in Figures 9 and 10 are very revealing and address any doubts that may be entertained about the behaviour of negative phase media.

Anisotropic and Photonic Crystals

Now that the discussion of causality has been brought to a successful conclusion, attention will be briefly turned to some other areas in which negative refraction is possible: crystal anisotropy and photonic crystals. The concept of negative refraction is so fresh and exciting in some areas that it may lay claim to being a new frontier in science. Statements surrounding the possible appearance of negative refraction can be exaggerated, however, and sometimes misleading; at best limited in scope. For example, anisotropic crystals, such as calcite, can exhibit amphoteric (from the Greek *amphoteros*, meaning each of two) refraction (Zhang et al. 2003) i.e. refraction can occur positively or negatively. The linkage of this to the general bandwagon associated with negative phase media is incorrect, though, because, as will be seen below, this is a forward wave, or positive phase, phenomenon.

The outcomes for beams or pulses interacting with crystal surfaces needs a full FDTD simulation, if detailed information is required. What the general conclusions show, however, is that a construction based upon rays that carry energy (to which Snell's law *cannot* be applied), and wave fronts that are normal to a wave vector $\mathbf{k}$ (to which Snell's law *must* be applied) can be used to gain information about the type of refraction taking place. This is an important conclusion because anisotropy effects are a considerable complication beyond isotropy. It is perhaps not too surprising then that negative refraction at the interface between two anisotropic crystals has been claimed as being the same kind of negative refraction (Zhang et al. 2003) that the backward wave community is talking about, however misleading this interpretation is. This is because of the general impression created that all negative refraction is connected to backward wave phenomena. On the contrary, this section will emphasise that negative refraction has always been with us in almost every day examples, involving *forward waves* in anisotropic crystals.

To address the issues it is always necessary to emphasise the distinction between rays and wave fronts (Saleh and Teich 1991). As previously stated, rays transport energy whilst fronts point in the direction of the phase velocity. This is illustrated by the familiar example of the kind of double refraction exhibited by e.g. a calcite crystal. Double refraction and anisotropy are, of course, inextricably linked and this leads to ordinary and extraordinary rays having orthogonal polarisations. If the dispersion equation for light propagation in a crystal is $\omega = f(\mathbf{k})$, where ω is the angular frequency and $\mathbf{k}=(k_x, k_y, k_z)$ is a wave vector, then what can be called a $\mathbf{k}$ surface is the surface over which ω is a constant. This surface is a very useful representation, not least because the direction of the normal to a tangent to this surface is $\nabla_{\mathbf{k}}\omega(\mathbf{k})$ and that this quantity is $\mathbf{v_g}$, the group velocity. The Poynting vector is proportional to $\mathbf{v_g}$, so the energy rays will point in the direction of the normals to the $\mathbf{k}$ surface, while the phase velocity will be parallel to $\mathbf{k}$. Uniaxial symmetry will be adopted and the $\mathbf{k}$ surfaces, or wave fronts, are spherical for the ordinary wave and non-spherical for the extraordinary wave. Both types have the optic axis as an axis of symmetry, so the inclination of this optic axis to any interface will be of crucial importance to the refraction process. For a uniaxial crystal the $\mathbf{k}$ surface of constant phase for the extraordinary wave is actually elliptic. For the ordinary wave it is spherical. Figs.11a and 11b show, for a given surface the optic axis of the crystal can be inclined and a projection of the $\mathbf{k}$ space can be drawn using the (k_x, k_y) plane, for example. Matching this projection to air across a flat boundary immediately shows the refraction possibilities.

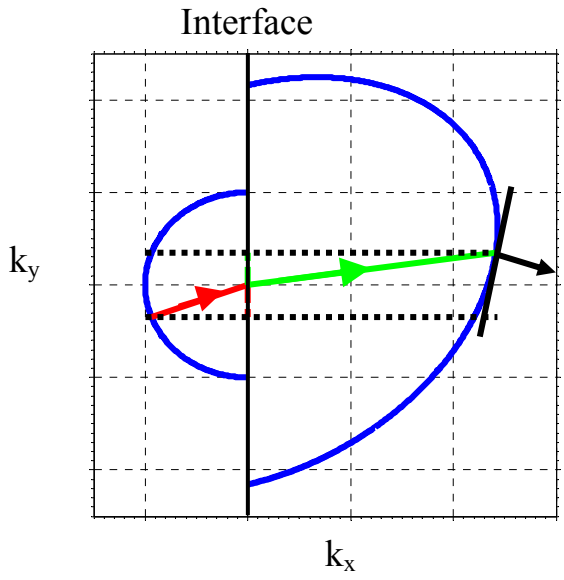


Figure 11a: Negative refraction at the interface between air and an anisotropic crystal. Note that the normal to the tangent shows the group velocity direction and $\mathbf{k}$ gives the phase velocity direction.

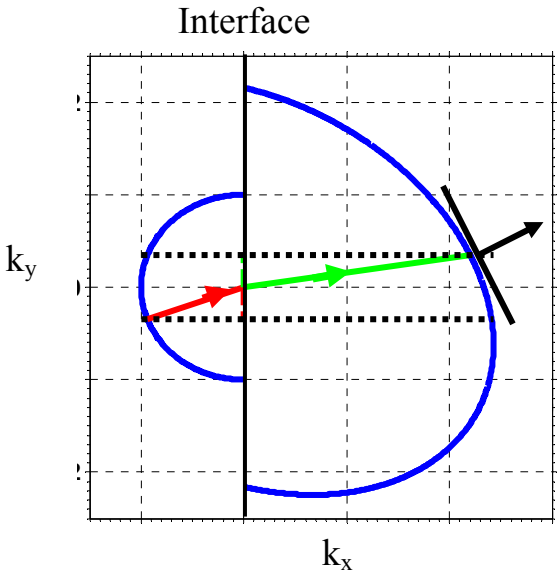


Figure 11b: Positive refraction at the same air/anisotropic interface.

For both the positive and negative refraction the phase waves are travelling forward but at an angle to the group velocity direction. Hence this is a *positive phase* medium, regardless of whether positive or negative refraction is taking place. Negative refraction in e.g. a calcite crystal is in fact a familiar property and shows up through the track of the extraordinary ray. It is very easily seen by simply handling a sample of calcite. Negative refraction in positive phase (right-handed) materials has always been with us then. The results from an FDTD simulation for the uniaxial crystal calomel are displayed in Figures 12 and 13.

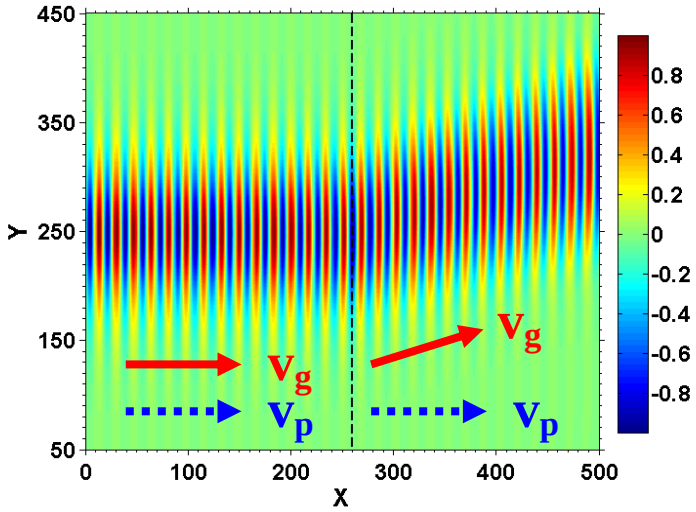


Figure 12: Generation of an extraordinary ray for a calomel crystal with the optic axis inclined at 37° to the y- axis. The beam is incident from air beam is normal incidence from air

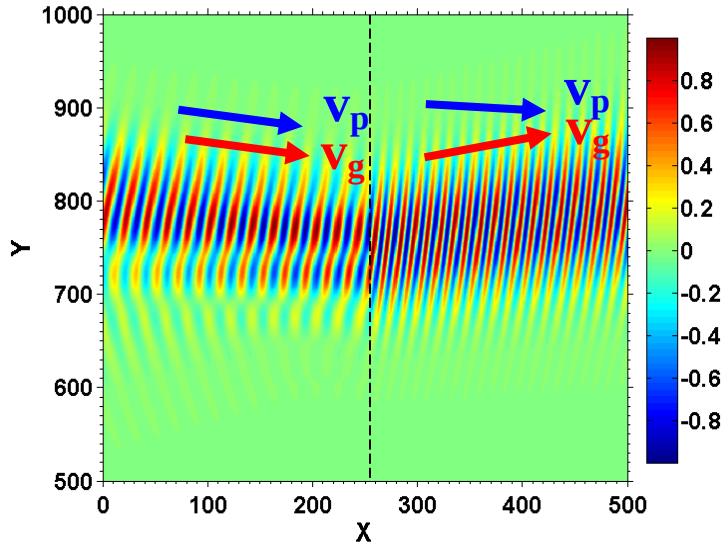


Figure 13: Extraordinary ray for a calomel crystal slightly inclined from the previous case of normal incidence to demonstrate the occurrence of negative refraction

These figures show that the negative refraction is accompanied by a *forward* travelling phase velocity but this phase velocity is moving at an angle to the group velocity. This positive phase refraction is quite different from the negative phase work that is currently catching such a lot of attention. Just as was discovered from Schuster's work, negative refraction is not a new idea, nor is it very surprising. Meanwhile the search for isotropic materials with backward wave properties gathers momentum and negative refraction continues to fascinate.

One area of fascination is the extent to which negative refraction can be generated by photonic crystals. Indeed, there is a strong desire for engineered materials exhibiting negative refraction at optical frequencies. Approaches to synthetic optical range negatively refracting media have led to the pursuit of negative phase metamaterials involving "nanostructures" (Podolskiy, Sarychev and Shalaev 2003; Grigorenko et al.

2005) but a photonic crystal solution is also very attractive (Foteinopoulou, Economou, and Soukoulis, 2003; Luo et al. 2002). Although photonic crystals are very interesting only their bearing upon the negative refraction perspective will be addressed here. It has been established that some features of photonic crystals produce negative refraction similar to that occurring in isotropic negative phase metamaterials. However, it appears that a degree of caution is needed because there is more than one mechanism for negative refraction in photonic crystals. This is rather different from the properties outlined above for the isotropic negative phase materials that lead directly to antiparallel group and phase velocities. In photonic crystals, negative refraction can occur even though it is still effectively a positive phase medium. Negative refraction can occur as a result of anisotropy, in a similar manner to the birefringent crystals discussed earlier. In other words, the gradient of an equifrequency contour in a photonic crystal need not be parallel to the wave vector, leading to forward wave negative refraction. In particular, for this type of photonic crystal operation, negative refraction will only occur under correct matching of the wave vectors and may not be allowed for certain directions through the crystal (Gralak, Enoch and Tayeb 2000). Also, positive phase negative refraction can occur due to coupling to a higher order Bragg wave (Foteinopoulou, Economou, and Soukoulis 2003). In spite of these concerns, it should be expected that the constant-frequency dispersion curves are circular around the centre of the first Brillouin zone. In that case, the usual negative phase behaviour may manifest itself if the frequency gradient is suitable. Even if this achieved, there is often residual anisotropy that can cause the Poynting vector and the wave vector to be non-collinear.

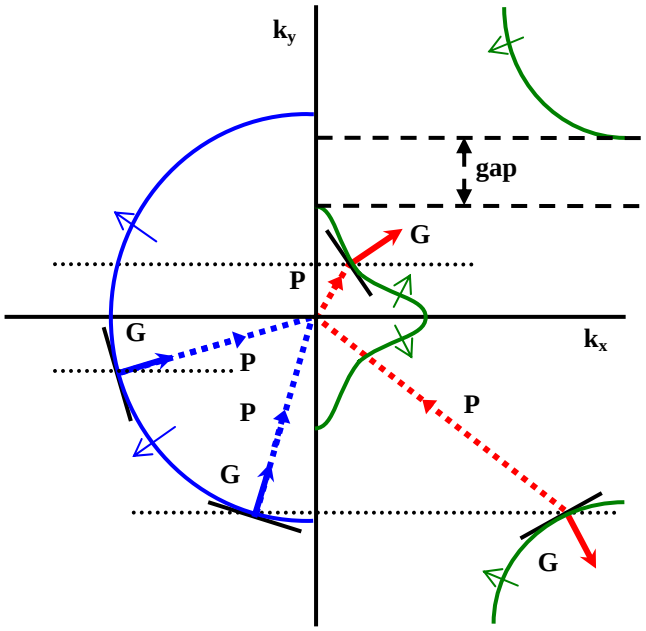


Figure 14: Diagram of wave number surfaces and group (G) and phase (P) wave directions for a possible photonic crystal interfaced to an isotropic positive phase medium. The directions of increasing gradient for the equifrequency contours are marked by unlabelled arrows. There is residual anisotropy causing non-parallelism.

This diagram shows two waves incident at different angles from a generic isotropic positive phase medium media onto a possible photonic crystal. In this case, the positive phase medium and the inner contour of the photonic crystal both have a

positive gradient, i.e. for increasing frequency these contours will sweep out towards increasing *k* values, and so the phase velocity is *forward* relative to the group velocity. However, the outer contour of the photonic crystal has a negative gradient and hence the phase velocity is *backward* relative to the group velocity. As well as indicating the anisotropic behaviour of photonic crystals, this diagram also demonstrates that, as the angle of incidence increased a gap (a region of no propagation) is encountered. Figure 14 looks complicated but it does in reality imply that negative refraction for this type of photonic crystal is a possibility as clarified below.

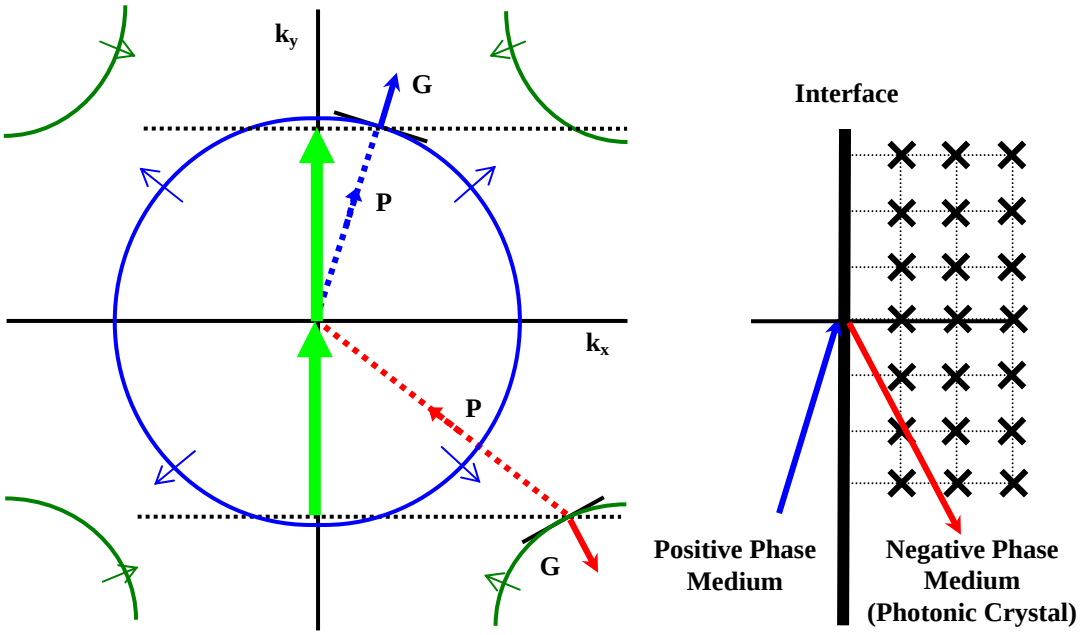


Figure 15: Extended phase matching construction to match isotropic medium to phase behaviour in a stylised square lattice photonic crystal. ✕ denotes a lattice site.

Figure 15 shows the usual phase matching construction mapped onto the ***k***-space expected from a simple square lattice selected to represent a typical photonic crystal. Only the outer frequency contours are used in this case because not only are they circular but they have the correct frequency gradient to guarantee the generation of negative refraction associated with negative phase velocity (Gralak, Enoch and Tayeb 2000). Note that absolute anti-parallelism can not be guaranteed and this is evidence of the residual anisotropy mentioned earlier in the text.

Conclusions

This paper provides a perspective on the concept of negative refraction. It is not intended as a comprehensive catalogue of all the papers published in this area. Its purpose is to address the concept of negative refraction and its historical and current background. Beginning with the ideas of Schuster, it is shown that the concept of negative refraction has been in the literature for a very long time but, nevertheless, in many areas of science, the excitement that surrounds this phenomenon is quite recent. It is emphasised that the modern search for negative refraction mainly involves metamaterials but it also can be found in a limited way in anisotropic crystals and, in quite a promising way, in photonic crystals. The latter could also be regarded as

metamaterials. The concept of forward and backward waves is clearly illustrated using FDTD simulations in which the direction of the phase front and the energy rays are defined through the nature of the method. It is demonstrated, visually that negative refraction with forward waves occurs in anisotropic crystals and that isotropic media can support backward wave negative refraction. It is hoped that this paper provides the perspective implied by the title.

References

Barjavel. R., (1973) *Le voyageur imprudent*. Paris: Editions Gallimard.

Boardman, A.D., ed.,(1982) *Electromagnetic surface modes.* New York: Wiley.

Boardman, A.D., Egan, P., Velasco, L. and King, N., (2005) Control of planar nonlinear guided waves and spatial solitons with a left-handed medium, *J. Opt. A: Pure Appl. Opt.* 7, pp. 40-60.

Foteinopoulou, S., Economou, E.N. and Soukoulis, C.M., (2003) Refraction in media with a negative refractive index, *Phys. Rev. Lett.* 90.

Gralak, B., Enoch, S. and Tayeb, G., (2000) Anomalous refractive properties of photonic crystals, *J. Opt. Soc. Am A.* 17, pp. 1012-1020.

Grigorenko, A. N., Geim, A. K., Gleeson, H. F., Zhang, Y., Firsov, A. A., Khrushchev, I. Y. and Petrovic, J., (2003) Nanofabricated media with negative permeability at visible frequencies, *Nature* 438.

Jaggard D.L., Mickelson A.R. and Papas C.H., (1979) On electromagnetic waves in chiral media. *Appl. Phys,* 18, pp. 211-216.

Kretschmann E., (1971) Determination of optical constants of metals by excitation of surface plasmons, *Z. Phys.* 241, pp. 313-324.

Lakhtakia, A., (2003) Handedness reversal of circular Bragg phenomenon due to negative real permittivity and permeability, *Opt. Exp.* 11, pp. 716-722.

Lakhtakia, A., (2003) On planewave remittances and Goos-Hänchen shifts of planar slabs with negative real permittivity and permeability, *Electromagnetics*. 23, pp. 71-75.

Lamb, H., (1916) *Hydrodynamics.* Cambridge: Cambridge University Press.

Lindell, I.V., Tretyakov, S.A., Nikoskinen, K.I. and Ilvonen, S., (2001) BW media – media with negative parameters, capable of supporting backward waves, *Micro. Opt. Tech Lett.* 31, pp. 129-133.

Luo, C., Johnson, S.G. Joannopoulos, J.D. and Pendry, J.B., (2002) All-angle negative refraction without negative effective index, *Phys. Rev. B*. 65.

Mandelshtam, L.I., (1945) Group velocity in a crystal lattice, Zh. Eksp. Teor. Fiz.(JETP) 15, pp. 475-478.

Maslovski, S.I., (2002) Negative refraction and causality, *URSI IEEE XXVII national convention on radio science*. Espoo, Innopoli.

McCall, M.W., Lakhtakia, A. and Weiglhofer, W.S., (2002) The negative index of refraction demystified, *Eur. J. Phys.* 23, pp. 353-359.

Pendry, J.B., (2003) Focusing light using negative refraction, *J. Phys.: Condens. Matter*. 15, pp. 6345-6364.

Pendry, J.B., (2004) Negative refraction, *Contemp. Phys.* 45, pp.191-202.

Pendry. J.B., Holden, A.J., Robbins, D.J. and Stewart, W.J., (1998) Low frequency plasmons in thin-wire structures, *J. Phys.: Condens. Matter*. 10, pp. 4785-4809.

Podolskiy, V.A., Sarychev, A.K. and Shalaev, V.M., (2003) Plasmon modes and negative refraction in metal nanowire composites, *Opt. Express*. 11: pp. 735-745.

Pokrovsky, A.L. and Efros, A.L., (2002) Sign of refractive index and group velocity in left-handed media, *Solid State Comms*. 124, pp. 283-287.

Ruppin, R., (2000) Surface polaritons of a left-handed medium, *Phys. Lett. A.* 277, pp. 61-64.
Saleh B.E.A. and Teich M.C., (1991) *Fundamentals of photonics*. New York: Wiley.
Schuster, A., (1904) *An introduction to the theory of optics.* London: Edward Arnold
Silin, R.A., (1978) Possibilty to produce plane-parallel lenses, *Opt. Spectrosc.* 44, pp. 89-191.
Smith, D.R. and Kroll, N., (2000) Negative refractive index in left-handed materials, *Phys. Rev. Lett.* 85, pp. 2933-2936.
Smith, D.R., Padilla, W.J., Vier, D.C., Nemat-Hasser, S.C. and Schultz, S., (2000) Composite medium with simultaneously negative permeability and permittivity, *Phys. Rev. Lett.* 84, pp. 4184-4187.
Smith, D.R., Schurig, D. and Pendry, J.B., (2002) Negative refraction of modulated electromagnetic waves, *Appl. Phys. Lett.* 81, pp. 2713-2715.
Taflove, A. & Hagness, S.C., (2000) *Computational Electrodynamics: The Finitie-Difference Time-Domain Method.* London: Artech House.
Toll J.S., (1956) Causality and the dispersion relation: logical foundations, *Phys. Rev.* 104, pp. 1760-1770.
Valanju, P.M., Walser, R.M. and Valanju, A.P., (2002) Wave refraction in negative media: always positive and very inhomogenous, *Phys. Rev. Lett.* 88, 187401.
Veselago, V.G., (1968) The electrodynamics of substances with simultaneously negative values of ε and μ, *Sov. Phys. Uspekhi*. 10, pp. 509-518.
Weiglhofer W.S. and Lakhtakia A. eds., (2003) *Introduction to Complex Mediums for Optics and Electromagnetics,* Bellingham, WA: SPIE Press.
Zhang, Y., Fluegel, B. and Mascarenhas, A., (2003) Total negative refraction in real crystals for ballistic electrons and light, *Phys. Rev. Lett.* 91,157404.
Ziolkowski, R.W., (2003) Pulsed and CW Gaussian beam interactions with double negative metamaterial slabs, *Opt. Exp.* 11, pp.662-681.
Ziolkowski, R.W. and Kipple, A.D., (2003) Causality and double-negative materials, *Phys. Rev. E*. 68, 026615.

Gender, Passive Stretch and Tendon Mechanical Properties

Katherine Burgess
Institute for Health and Social Care Research

Introduction

Humans develop force through the contraction of muscle, which is transmitted via tendon to bone causing movement or action against resistance. In order to generate external force muscles must first overcome the slack in the tendon (Wilson et al. 1994). As tendon stiffness increases force transmission from muscle to bone will be more rapid. This has been demonstrated practically by Bojsen-Moller et al. (2005) who found tendon stiffness to be related to rate of force development and electromechanical delay.

The use of stretching exercises to improve range of motion is common practice among both recreational and competitive athletes (Kubo et al. 2002). However, stretching has be shown to acutely decrease performance in; one repetition maximum exercises (Kokkonen et al. 1998; Nelson and Kokkonen 2001), various jump exercises (Young and Elliott 2001), maximal isokinetic efforts (Marek et al. 2005) and sprint performance (Nelson et al. 2005). Evetovich et al. (2003) reported a decrease in muscle stiffness as indicated by an increase in mechanomyography following static stretching. Similarly Magnusson et al. (1996) showed a decrease in muscle stiffness after repeated static stretching, stiffness was defined as the change in torque (in Newton-meters) divided by angular change in position (in radians) and was expressed as the slope of the torque-position curve, (measurements were made during the dynamic phase of a passive stretch). In addition to muscle stiffness recent methodological advances have enabled tendon stiffness to be measured in vivo using ultrasonography. Using this method Kubo et al. (2001) found gastrocnemius tendon stiffness in males to decrease significantly following static stretching. This decrease in muscle/tendon stiffness has been proposed as a mechanism for diminished performance following stretch (Cornwell et al. 2002).

Kubo et al. (2003) showed that there are gender differences in the viscoelastic properties of tendon structures. They found that the stiffness and Young's modulus of the gastrocnemius tendon was significantly lower in women (16.5 ± 3.4 N/mm, 277 ± 25 MPa) than in men (25.9 ± 7.0 N/mm, 356 ± 32 MPa). These observations may partly explain the difference in measures of performance between genders, especially those which are dependant on force generation in the early stages of muscle contraction such as electromechanical delay and rate of force development.

As tendon mechanical properties between genders have been shown to be different, it may be that there are differences in the effect an acute stretch has on the structure of the tendon. By its nature if a set force is applied to a tendon which is more compliant it will stretch further and so experience a greater strain than a stiffer tendon. It is therefore suggested that stretching will have a greater effect in terms of changes in stiffness in females whose tendons are initially more compliant that males.

Therefore, the present study aims to compare the effects of stretching on tendon stiffness between genders.

Method

Subjects

Sixteen males, mean age 22 years ± 2 (SD), height 181.1cm ± 6.6 and body mass 83.4kg ± 11.4, and 10 females, mean age 20 years ± 1 (SD), height 166.0cm ± 7.1 and body mass 65.3kg ± 10.4 participated in the study. Participants were recruited from the university population and were all habitually active; all females were currently taking an oral contraceptive which contained oestrogen. Prior to selection all subjects were screened for evidence of previous lower limb injury and any other medical condition which could have prevented them from participating. The investigation was approved by the Salford University Institutional Ethics Committee and all subjects gave their written informed consent to participate in the experiment. The study conformed to the principles of the World Medical Associations Declaration of Helsinki.

Experimental Design

The mechanical properties of the medial gastrocnemius tendon, in both males and females, were examined prior to and immediately post completion of a passive dorsiflexion stretch.

Measurement of Torque

A dynamometer (Kin Kom) was used to measure the torque output during isometric plantar flexion. The participants were seated on the dynamometer with the knee fully extended and the hip flexed to 90^0. The foot was fixed in a neutral anatomical position, where the sole of the foot was at 90^0 to the tibia. The ankle joint axis was visually aligned with the pivot point of the dynamometer leaver and the foot was securely fixed to the dynamometer foot plate with Velcro straps. Prior to the test, the participants performed three warm up isometric planter flexions in order to accustomise themselves to the procedure. The participants were instructed to gradually develop force from relaxed to maximum voluntary contraction (MVC) over a 10s time period. The task was repeated three times with a 1 minute rest between trials. The torque signals were analogue-to-digital converted at a sampling rate of 2 KHz (Testpoint, Keithley instruments, UK) and saved for further analysis.

Measurement of Tendon Elongation

Tendon elongation measurements were taken during the graded isometric plantar flexion test using a 7.5MHz, 40mm linear array, B-Mode ultrasound probe (AU5, ESAOTE BIOMEDICA, Italy) with a depth resolution of 49.3mm. The probe was placed in the sagittal plane over the myotendinous junction of the medial head of the gastrocnemius muscle and fixed in position. An echo-absorptive marker was placed between the probe and the skin to act as a fixed reference from which measures of elongation could be made. The ultrasound image was displayed in real time on the ultrasound monitor. The S-VHS output video signal from the ultrasound apparatus was fed to a computer and captured at 25Hz using Adobe Premier Pro. The force output from the dynamometer and the ultrasound video were synchronised using a trigger (Digitimer, UK) that provided a visual marker on the ultrasound image and force trace. Stills from the ultrasound video were taken at times corresponding 10% force increments from 0 to 100% MVC and the displacement of the gastrocnemius

myotendinous junction were digitised using computerised image analysis (Image J, Wayne Rasband National Institute of Health, USA).

Measurement of Electromyographic Activity

Two silver/silver chloride bipolar electrodes (Medicotest UK, type N10A), with a 20mm inter-electrode distance (centre to centre) were placed midline on the muscle belly, halfway between the centre of the belly and the distal myotendinous junction of the tibialis anterior. A ground electrode (Medicotest, UK, type Q10A) was placed at an electrical neutral site (lateral malleolus of the ankle). Electrodes were aligned perpendicular to the direction of muscle fibres. The participants' skin was carefully prepared through a process of shaving, abrasion (Nuprep, SLE Ltd) and cleaning with alcohol before the electrodes were attached in order to minimise resistance. The Electromyographic signals were high and low pass filtered between 10 and 500 Hz respectively (Neurolog filters NL 144 and NL 134, Digitimer, UK), preamplified (x1000), (Neurolog remote AC preamplifier NL 824, Digitimer, UK), amplified (x2) (Neurolog isolation amplifier, NL 820, Digitimer, UK) and A/D converted at a rate of 2000Hz (KPCI 3101, Keithley instruments, UK). Post acquisition, a moving root mean square filter (100 ms window) was used to filter the electromgogram (EMG) signal.

In addition to the plantar flexion efforts participants performed three dorsi flexion MVC's. The dorsiflexions were performed with the subject in the same position as previously described for the plantar flexion efforts. The force produced due to dorsiflexor muscles coactivation during the plantar flexion efforts was approximated assuming a linear relationship between EMG amplitude of the dorsiflexor muscles and force.

$$\text{Co-contraction Torque} = \frac{\text{TibAnt EMG during Plantarflexion}}{\text{TibAnt EMG during Dorsiflexion}} \times \text{Maximal Dorsiflexion Torque}$$

Moment Arm Determination

The moment arm length of the gastrocnemius tendon was obtained using the tendon travel method (An et al. 1984). The displacement of the myotendinous junction caused by rotating the ankle from 5^0 of dorsiflexion to 5^0 of plantarflexion was recorded using ultrasonography (using the same probe and positioning described above in the measurement of elongation). The tendon moment arm length at the ankle angle of 0^0 was obtained from the ratio of change in tendon displacement in mm to change in angle in radians.

Calculation of Tendon Force

The torque values obtained from the dynamometer were converted into forces experienced by the tendon by dividing by the length of the tendon moment arm as measured using the tendon travel method explained above. Correction for relative muscle physiological volume of the gastrocnemius was applied in the calculation of final forces as per Fukunaga et al. (1996). Antagonistic co - contraction torque was determined and corrected for during the measures using sEMG (see 'measurement of electromyographic activity' section).

Calculation of Tendon Stiffness

The elongation of the tendon at loads corresponding to 0-100% of the plantarflexion force produced was measured at 10% intervals. The force elongation relationship was plotted and a second order polynomial fit was applied. Tendon stiffness was

calculated at 10% intervals of MVC and was defined as the slope of the force displacement relationship.

Stretch

A five minute passive dorsiflexion stretch was administered to each participant. The participant was positioned in the dynamometer as in the plantar and dorsi flexion efforts (knee straight, hip at 90^0). The participants were instructed to remain as relaxed as possible during the stretch and to ensure this was the case the EMG activity of the medial gastrocnemious and tibialis anterior was monitored (see measurement of electromyographic activity section for details). The foot plate of the dynamometer (to which the participants foot was attached) was moved into dorsiflexion so that it recorded a passive torque of 35-40Nm above baseline, this torque level was monitored and maintained throughout the stretch.

Statistics

Descriptive data included means (SEM). The significance of the difference between before and after passive stretching was analyzed by a paired Student's t-test. The level of significance was set to $P<0.05$.

Preliminary Results

Females were shown to have smaller tendon moment arms (46.78 ± 5.3mm) than males (53.42 ± 2.6mm).

Medial gastrocnemius tendon stiffness (at 100% MVC) prior to stretching was greater in males (68.9 ± 7.8 Nmm^{-1}) than females (53.5 ± 8.5 Nmm^{-1}). With female's stiffness being equal to 77.6% of males.

Post stretching there was no significant change in male's tendon stiffness (68.3 ± 7.1 Nmm^{-1}) whereas females showed a 26.42% decrease in stiffness ($p<0.05$) to a value of 39.2 ± 6.1 Nmm^{-1} (see figure 1).

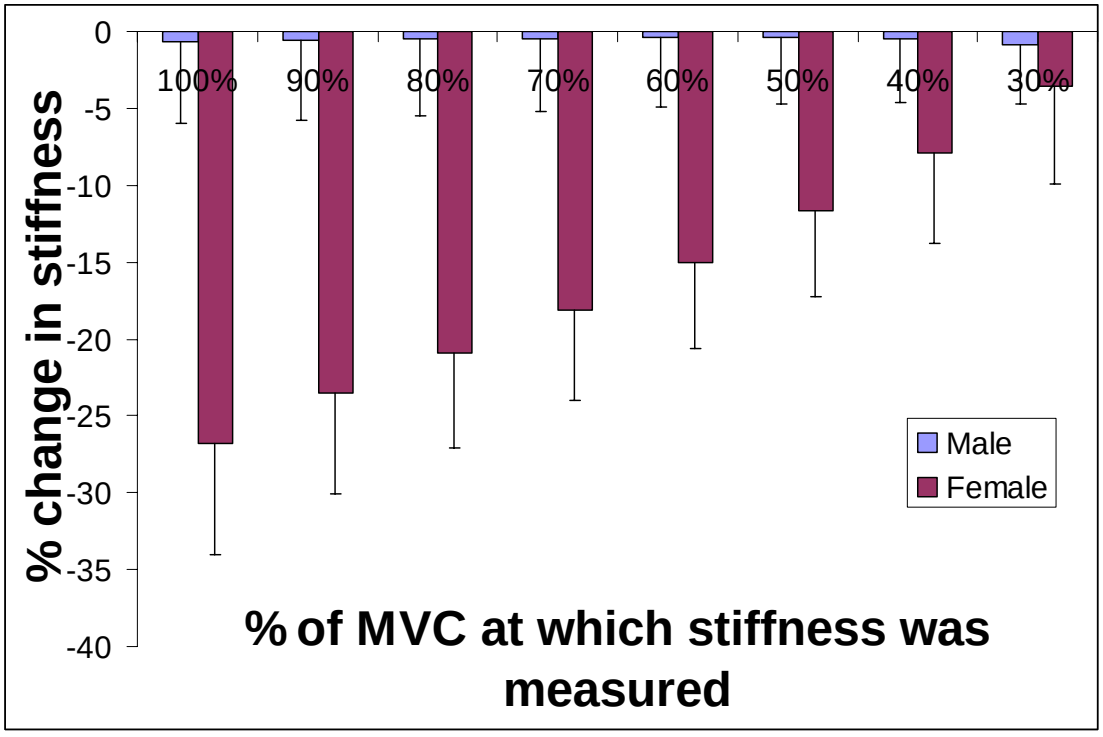


Figure 1: Change in gastrocnemius tendon stiffness for males and females at percentages of MVC

Maximum male tendon displacement increased from 15.39mm to 15.81mm post stretching (see figure 2).

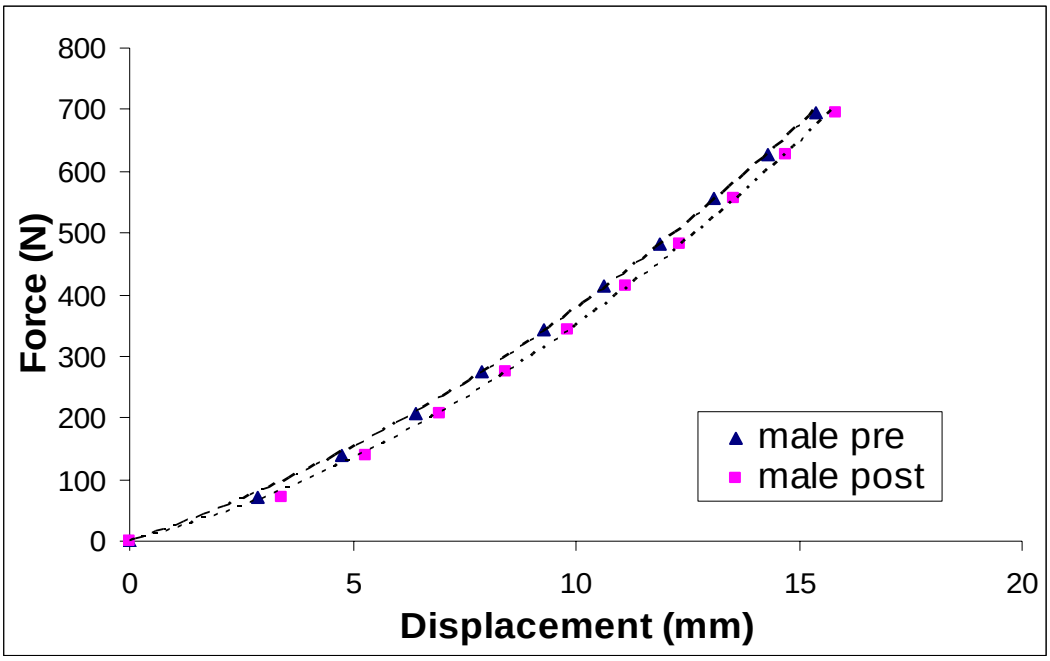


Figure 2: Male force elongation profile pre and post stretching

Maximum female tendon displacement increased from 11.37mm to 12.46mm post stretching (see figure 3).

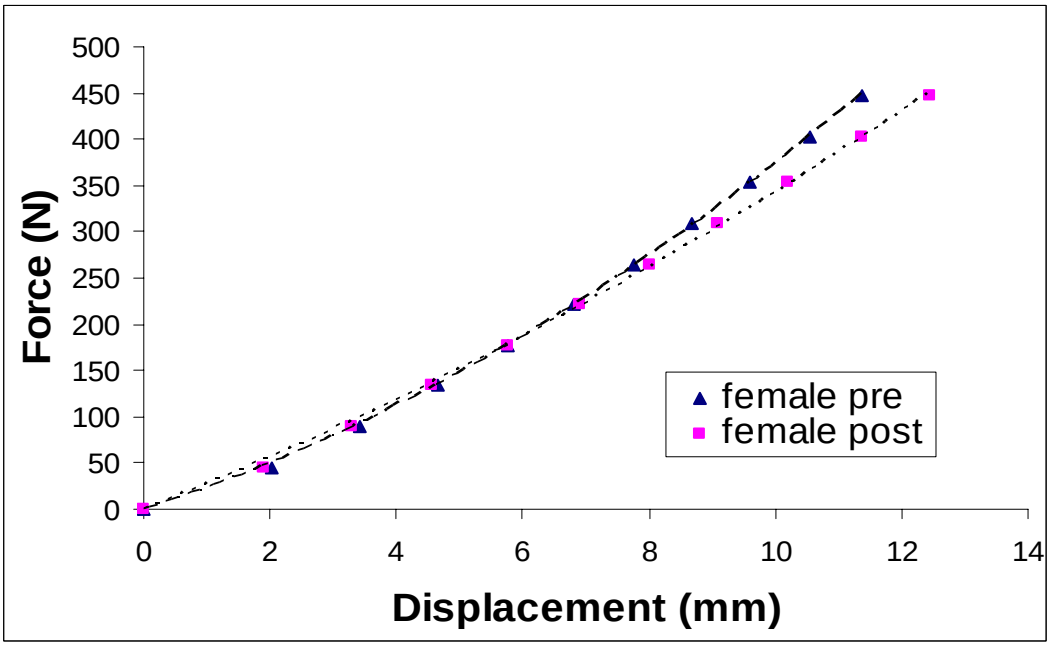


Figure 3: Female force elongation profile pre and post stretching

Discussion

The average tendon stiffness prior to the performance of the passive stretch was 68.9 ± 7.8 Nmm^{-1} for males and 53.5 ± 8.5 Nmm^{-1} for females. These values lie within the

range of values obtained by previous experiments which have used ultrasonography to measure gastrocnemius tendon stiffness in vitro. Kubo et al. (2003) reported lower values of 25.9 ± 7.0 Nmm^{-1} for males and 16.5 ± 3.4 Nmm^{-1} for females whereas Maganaris and Paul (2002) reported larger values of 149.8 Nmm^{-1}. These differences are probably due to variance in the population group utilised, differences in the methods utilised to determine tendon force and displacements and differences in the methods used to calculate tendon stiffness.

Several studies (including those stated above) have not accounted for co-contraction causing underestimations in force. Different methods are used to determine tendon moment arm, some studies use magnetic resonance imaging (Magnusson et al., 2001), others use the tendon travel method which was utilised in this study (Maganaris and Paul 2002) and Kubo's group (Kubo et al. 2002; Kubo et al. 2001) utilise a method based on a study by Visser et al. (1990) which determines the moment arm as a function of leg length, however the validity of this method is questionable. In order to estimate the force solely produced by the medial gastrocnemius from plantar flexion force, the most commonly used method is that of volume or cross-sectional area correction ratio. The assumptions made with this method can lead to some inherent error as not all subjects will have the same muscle volume/ cross-sectional area proportions. However, in the present study we aimed to examine whether tendon properties altered after stretching and so force calculations based on this assumption would be valid to study the changes within individuals. Studies which have not accounted for this fraction have reported inflated stiffness values due to inaccurately high force values such as 486 Nmm^{-1} (Magnusson et al. 2001) and 306 Nmm^{-1} (Rosager et al. 2002).

As previously shown by Kubo et al. (2003) these results show that female's gastrocnemius tendon is more compliant than males with females' stiffness being 77.6% of males compared to 63.7% reported by Kubo.

Post stretching there was a significant decrease in tendon stiffness in females of -26.42 ± 7.2% ($p<0.05$) whereas there was no significant change in males (0.99 ± 5.2%). Previous research using ultrosonography to measure tendon stiffness prior to and post stretching has been carried out on males and has produced contradictory findings to that of this study. Kubo et al. (2001) reported a 10% decrease in stiffness following a 10minute passive dorsiflexion stretch held at 35^0 and similarly Kubo et al. (2002) showed a 7.9% decrease in stiffness following the same stretch held for 5 minutes. The stretch protocol used in this study was similar to that used by Kubo et al. (2002) in that the foot was held in a passive dorsiflexion stretch, however, instead of setting the angle of stretch, the stretch was held in position so that the limb was experiencing a passive torque of 35-40Nm above baseline throughout the stretch. These forces were reported as being the peak passive force at the onset of stretch by Kubo et al. (2002). Ensuring these passive forces remained constant and did not decrease due to stress relaxation meant the severity of the stretch was maintained and so this difference could only led to a harsher stretch being applied and thus less of a change would be unexpected based on methodological differences. The values of tendon stiffness reported by Kubo et al.,(2002) prior to stretch are lower than those reported in this study, the differences caused by the stretch could well be due to differences in initial compliance.

This adds weight to our finding that females (whose tendons are originally more compliant) decreased their tendon stiffness significantly following the same stretch. During the stretch the same external load was applied to the system in terms of passive force, however, for this same level of passive force a more compliant tendon will experience a greater degree of elongation and therefore strain. This greater strain could cause a larger change in the tendon properties or structure.

The mechanisms which cause an increase in tendon compliance following stretching are largely unknown, changes in the structure of the tendons in terms of length and cross-sectional area may well be involved and will be monitored during the next stage of this study. In unstressed tendons the collagen fibres follow a wave like course, however when stressed these have been shown to become more aligned (Stromberg and Wiederhielm 1969). This may well occur due to stretch and could effect the tendons properties.

Conclusion

The present study confirmed previous reports that there are differences in tendon properties in terms of stiffness between genders. Furthermore, passive stretching significantly decreases tendon stiffness in females but has no significant effect in males. This difference in adaptation is thought to be due to the variance in tendon structure and/or properties between genders, more compliant structures showing a greater change in stiffness following stretch.

References

An, K. N., Takahashi, K., Harrigan, T. P. and Chao, E. Y., (1984) Determination of Muscle Orientations and Moment Arms, *Journal of Biomechanical Engineering-Transactions of the Asme,* 106, pp. 280-282.

Bojsen-Moller, J., Magnusson, S. P., Rasmussen, L. R., Kjaer, M. and Aagaard, P., (2005) Muscle performance during maximal isometric and dynamic contractions is influenced by the stiffness of the tendinous structures, *Journal of Applied Physiology,* 99**,** pp. 986-994.

Cornwell, A., Nelson, A. G. and Sidaway, B., (2002) Acute effects of stretching on the neuromechanical properties of the triceps surae muscle complex, *European Journal of Applied Physiology,* 86**,** pp. 428-434.

Evetovich,, T. K., Nauman, N. J., Conley, D. S. and Todd, J. B., (2003) Effect of static stretching of the biceps brachii on torque, electromyography, and mechanomyography during concentric isokinetic muscle actions, *Journal of Strength and Conditioning Research,* 17**,** pp. 484-488.

Fukunaga, T., Ito, M., Ichinose, Y., Kuno, S., Kawakami, Y. & Fukashiro, S., (1996) Tendinous movement of a human muscle during voluntary contractions determined by real-time ultrasonography, *Journal of Applied Physiology,* 81**,** pp. 1430-1433.

Kokkonen, J., Nelson, A. G. and Cornwell, A., (1998) Acute muscle stretching inhibits maximal strength performance, *Research Quarterly for Exercise and Sport,* 69**,** pp. 411-415.

Kubo, K., Kanehisa, H. and Fukunaga, T., (2002) Effects of transient muscle contractions and stretching on the tendon structures in vivo, *Acta Physiologica Scandinavica,* 175**,** pp. 157-164.

Kubo, K., Kanehisa, H. and Fukunaga, T., (2003) Gender differences in the viscoelastic properties of tendon structures, *European Journal of Applied Physiology,* 88**,** pp. 520-526.

Kubo, K., Kanehisa, H., Kawakami, Y. and Fukunaga, T., (2001) Influence of static stretching on viscoelastic properties of human tendon structures in vivo, *Journal of Applied Physiology,* 90**,** pp. 520-527.
Maganris, C. N. and Paul, J. P., (2002) Tensile properties of the in vivo human gastrocnemius tendon, *Journal of Biomechanics,* 35**,** pp. 1639-1646.
Magnusson, S. P., Aagard, P., Rosager, S., Dyhre-Poulson, P. and Kjaer, M., (2001) Load-displacement properties of the human triceps surae aponeurosis in vivo, *Journal of Physiology-London,* 531**,** pp. 277-288.
Magnusson, S. P., Simonenson, E. B., Aagarrd, P. and Kjaer, M., (1996) Biomechanical responses to repeated stretches in human hamstring muscle in vivo, *American Journal of Sports Medicine,* 24**,** pp. 622-628.
Marek, S. M., Cramer, J. T., Fincher, A. L., Massey, L. L., Dangelmaier, S. M., Purkayastha, S., Fitz, K. A. and Culbertson, J. Y., (2005) Acute effects of static and proprioceptive neuromuscular facilitation stretching on muscle strength and power output, *Journal of Athletic Training,* 40**,** pp. 94-103.
Nelson, A. G., Driscoll, N. M., Landin, D. K., Young, M. A. and Schexnayder, I. C., (2005) Acute effects of passive muscle stretching on sprint performance, *Journal of Sports Sciences,* 23**,** pp. 449-454.
Nelson, A. G. and Kokkonen, J., (2001) Acute ballistic muscle stretching inhibits maximal strength performance, *Research Quarterly for Exercise and Sport,* 72**,** pp. 415-419.
Rosager, S., Aagaard, P., Dyhre-Poulsen, P., Neergaard, K., Kjaer, M. and Magnusson, S. P., (2002) Load-displacement properties of the human triceps surae aponeurosis and tendon in runners and non-runners, *Scandinavian Journal of Medicine & Science in Sports,* 12**,** pp. 90-98.
Stromberg, D. D. and Wiederhielm, C. A., (1969) Viscoelastic description of a collagenous tissue in simple elongation, *J Appl Physiol,* 26**,** pp. 857-862.
Visser, J. J., Hoogkamer, J. E., Bobbert, M. F. and Huijing, P. A., (1990) Length and Moment Arm of Human Leg Muscles as a Function of Knee and Hip-Joint Angles, *European Journal of Applied Physiology and Occupational Physiology,* 61**,** pp. 453-460.
Wilson, G. J., Young, W. B., Hurnall, G., Hood, K. P. and Steinbronn, M., (1994) The effect of slow versus fast rate of force development on resistance training induced changes in performance, *Journal of Human Movement Studies,* 27**, pp.** 75-88.
Young, W. and Elliot, S., (2001) Acute effects of static stretching, proprioceptive neuromuscular facilitation stretching, and maximum voluntary contractions on explosive force production and jumping performance, *Research Quarterly for Exercise and Sport,* 72**,** pp. 273-279.

The Discourse on Sustainable Energy Futures – Three Story Lines

Roman Buss

Institute for Social, Cultural and Policy Research

Introduction

Over thirty years after the first oil crisis, energy is back on top of political agendas around the world. Although industrialised countries since the 1970s oil crisis have decoupled their economic growth from energy intensity, most countries and regions have not significantly reduced their dependence on fossil fuels. Energy-related greenhouse gas (GHG) emissions are widely seen as cause for global warming. Energy technologies are one of the main sources of many atmospheric environmental problems. Emissions in the power, transport, and domestic sectors remain on a high level and rarely display signs of significant reductions or even show constant increases. Some developments suggest that we are rapidly approaching the point where demand overruns the level of supply capacities with uncertain outcomes for the global economy, which is also reflected in the growing 'end-of-oil' literature (see Lovins et al. 2005; Roberts 2005; DB Research 2004).[1] Record fuel, oil and energy prices that have tripled during the last years also boosted calls for alternative energy sources. The war in Iraq, rapid oil and coal demand increases in developing economies in Asia, political turmoil in oil exporting countries like Iran, Nigeria or Venezuela, gas supply interruptions from Russia, hurricanes, oil depot blasts, and electricity blackouts in recent months have made policy makers and the public once more aware of the vulnerability of energy systems in an ever more turbulent world. It is before this background that policy-makers and scientists start to reconsider new configurations of global and national energy systems. At first sight there is a consensus among scientific communities and decision-makers in energy policy about the following assumptions:

> 1.) Levels of energy-related green house gas emissions are a threat to human development and the global climate. In order to stabilise atmospheric CO_2 concentrations and to avoid irreversible damage to ecosystems, rigorous reductions of around 60 per cent will have to be realised by 2050.
> 2.) World energy demand is expected to increase substantially during the next decades.
> 3.) Considerable parts of the ageing power and energy infrastructure in Western world will have to be modernised or replaced during the next decades.
> 4.) Modern economies are over-dependent on finite fossil energy sources, which are unequally distributed and often situated in politically unstable regions
> 5.) One fifth of the world's population consumes 80 per cent of all energy

In other words: Today's energy systems are not sustainable in the sense that they reconcile economic, social, and environmental concerns and aspects of intra-

[1] See also The end of oil is closer than you think, *The Guardian*, April 21, 2005.

generational and inter-generational justice. The main challenge of energy policy therefore is to meet a growing energy demand in a *more sustainable* way. Energy policy in general and specifically sustainable energy presents a considerable problem for analysis in that we find a multi-layered and interdisciplinary complex of interdependent problem dimensions and interest structures. On the surface there is agreement that a transition towards cleaner or sustainable energy futures has to be initiated.[2] But by scratching this surface we open a Pandora's box of differing notions of sustainable energy and means of achieving it. The catch phrase of *sustainable energy* often obscures deep-rooted conflicts in energy policy, as soon as the leitmotif is to be translated into concrete measures and policies.

Discourses in Environmental Policy and Sustainable Development

Following the linguistic or pragmatic turn in philosophy social science and policy analysis in the last decade have approached international environmental policy from the angle of discourses and conceptualised the ecological crisis as discursive constructions (MacRae 1993; Roe 1994. Fischer and Forester, eds. 1999; Hajer 2000; Feindt and Oels 2005; Hajer and Versteeg 2005). According to Hajer (2000) social constructivism and discursive practices of policy analysis in environmental social science emerged in the context of a wider post-positivist, interpretative, and argumentative tradition that argues that power structures of a society can and should be studied directly through discourse. Although modernist, linguistic, or critical versions can be distinguished the metaphysical paradigm of social constructivism relies on the assumption "that the objects of our knowledge are not independent of our interpretations and our language". Unlike Positivism and materialism which take the world as it is, social constructivism sees the world as becoming rather than being'(Adler 2002: 95). In this way environmental conflict and policy have also changed and become discursive. Disputes are no longer about whether there is an ecological crisis but about its interpretation. "Political conflict is hidden in the question of what definition is given to the problem, which aspects of social reality are included and which are left un-discussed" (Hajer 2000: 43). This approach is specifically helpful in complex issue areas like environmental policy or climate change where the public is confronted with "a wide range of views, with experts and counter-experts, with debates among scientists from different disciplines or different countries, and realise that scientific controversy is an inherent element of environmental politics" (Hajer 2000: 11).Central and common to these approaches is that language does not only reflect and mirror the world but also profoundly shapes our view of reality and the importance of narratives, meta-narratives, and story lines in the policy formation process. Hajer and Versteeg (2005) have defined these discourses as "ensemble of ideas, concepts and categories through which meaning is given to social and physical phenomena. . . Actors position themselves in the realm of given discourses and try to shape discourses". A new terrain of discourse evolves when experts (or parts of the expert community) define and attach urgency to a specific problem. Problem framings are based and rely on story lines, interest structures and value assumptions or core beliefs and are very resilient to change. This leads to symbolic struggles of dominance in the discourse in that actors also have to address the opponent's assumptions and beliefs (Brand and Jochum 2000: 173). The politics of sustainable development thus is not the product of a "linear, progressive,

[2] For discussions of transformations or an energy turnaround see for example Matthes/Cames (2002), WBGU (2003), Bechberger/Reiche (2006).

and value-free process of convincing actors of the importance of the green case". Politics is this view is conceived of as a struggle between unconventional 'discourse coalitions' (Hajer) or advocacy coalitions (Sabatier (1993); Heritier (1993); Sabatier and Jenkins-Smith (1999) for discursive hegemony in which actors try to secure support for their definition of reality.[3]

> Discourse–coalitions are defined as ensemble of (1) a set of story-lines; (2) the actors who utter these story-lines; and (3) the practices in which this discursive activity is based. Story lines are here seen as the discursive cement that keeps a discourse-coalition together. . . The discourse coalitions then consist of such actors as scientists, politicians, activists, or organizations representing such actors, but also having links with specific television channels, journals and newspapers, or even celebrities. . . They are distinct from each other in that they develop and sustain a particular discourse, a particular way of talking and thinking about environmental politics (Hajer 2000: 12-3).

The Sustainable Energy Discourse

This paper attempts to map out the rhetoric of *'sustainable energy'*. This is a good example for a contested discourse and struggle for discursive hegemony, which at the same time is a political 'battle over institutions' and a scientific competition between technological pathways (Martin 1996; Jacobsson/Laube 2006: 257). It is argued here that three discourse coalitions with largely exclusive belief systems three particular ways of talking and thinking about sustainable energy futures can be identified. The story lines *End-of-Pipe (EOP), Continuity* and *Discontinuity* can be understood as ideal-type policy responses to the problems of climate change.[4] These narratives stated by various actors in energy policy and businesses are often presented as authoritative truths and are sometimes accompanied by a religious-like aura of inevitability.

All three story lines or belief systems claim to arrive at a 'more sustainable' energy system. But opinions widely differ in terms of basic assumptions, use of political and economic instruments, chosen technologies, priorities of energy sources, the prescribed role of the state, timescales of transitions and relevant actors involved in the process. *End-of-Pipe* tackles emissions at the end of the energy supply chain, *Continuity* modifies selected components of the energy system with the means of moderate modernization, and *Discontinuity* represents a radical system change. The future direction of energy policy depends on which of those belief systems will achieve discourse institutionalization and translate its claims into policies.

The END-OF-PIPE (EOP) Narrative

The EOP narrative is best characterised as technocratic, large-scale, centralised, and industrialised model of energy policy and economics and as modified business-as-usual scenario of energy futures. It takes over its name from the approach that modifies

[3] Note that Hajer distinguishes his argumentative discourse coalition model from that of Sabatier's advocacy coalition framework. "Story lines, not interests, form the basis of the coalition. . . There are three major points where Sabatier's advocacy coalitions can be seen to be essentially different from discourse coalitions: (1) the individualist ontology differs from my relational ontology; (2) the central role of beliefs in advocacy-coalitions differs from my emphasis on the constitutive role of language and the role of story lines and discursive affinities; (3) Sabatier's notion of policy-oriented learning differs from my theory of social change" (Hajer 2000: 68-72).

[4] These terms have been taken over from Gregory Unruh's work on carbon lock-in. See (Unruh 2000, 2002). The analytical categories will be transferred to the discourses around sustainable energy.

energy infrastructures and technologies at the end of the technological supply chain.[5] The entire line of argument rests upon the assumption that fossil energy sources and nuclear power are here to stay and constituting an essential part of a 'clean energy future'. This discourse does not aim at a fundamental system change but argues that a sustainable energy system is the result of technological improvements of the existing energy facilities and infrastructures. The transition itself is seen as a slow and gradual process over the next four or five decades. The statistical range of fossil resources is estimated far into the 21st century. In the EOP narrative energy security and energy independence ('domestic sources of energy') but not primarily environmental concerns become major drivers of change. Any change in the energy structure of an energy system has to rely on the expertise and participation (consultation) of energy companies and utilities. In a subtle way it is hinted at the stability of energy supply in the past, suggesting that too sudden changes pose a threat to energy security. The massive de-carbonization of economies is interpreted as sufficient condition to qualify it as sustainable energy system. EOP stresses the "inescapable reality that hydrocarbons are going to power the vast majority of our vehicles, and generate the largest share of the world's electricity, for decades to come".[6] As the transformation to sustainable energy markets will take a considerable time, fossil fuel-based energy markets have to be transformed to mitigate the impact of the use of fossil fuels in the meantime (Gosselink 2002). EOP in energy policy and development thus aims at the winding-down of fossil energy use and emissions rather than the start of a deep transition process. The only bottlenecks in this scenario are climate related CO_2 emissions, but not the depletion of resources. EOP is therefore looking at ways to redesign carbon-burning systems to ensure that less or no carbon is released. Inherent to the EOP narrative is an engineering and technocratic notion of large-scale (and yet to be developed) technologies.[7] An ideal type EOP position was taken in a report by the Geological Society in the UK (2005). It emphasises that

> fossil fuels will remain our most important source for the next 50 years. Clean systems, including carbon capture and storage, should be pursued urgently; Nuclear fission energy is a proven and reliable technology that will inevitably have a key role in a future clean energy mix; Renewable energy sources will play a growing role, but will require continued support in development and deployment if they are to match the cost levels of conventional systems.[8]

[5] The term 'end-of-pipe' itself may be an unfortunate choice, because it has negative associations with perceptions prevalent during the 1970s when environmental policy saw environmental externalities as a necessary evil that accompanies economic activity and that can be technologically dealt with. See Janicke et al. (1999). Measures to combat environmental damage were placed at the end of production and consumption patterns (for example, filter technologies in power stations, catalysts in cars), but in the process itself. Yet I decided to use these terms, because they also reflect basic features of energy supply chains, from primary energy source, conversion, distribution to final end-use. End-of-pipe visualises best the notion of energy chains that are modified at the end.

[6] Has Blair the Energy to Convince Exxon?, *The Guardian*, 15 September 2004.

[7] For a comparison of renewable technologies and carbon capture and sequestration see Wuppertal Institute (2005).

[8] Geological Society (2005): How to plug the energy gap, (www.geolsoc.org.uk), 10 November 2005. Another archetypical example for the 'big technology' perspective is the energy scenario study by the German Physical Society (2005). In order to meet emission targets the DPG argues for lifetime extension of nuclear power plants and for large-scale solar-thermal installations in the Southern hemisphere.

EOP Perceptions of Fossil, Nuclear, and Renewable Energies

One of the main assumption of the EOP narrative is, that, if we like it or not, the fossils aren't dead and they can make a contribution to reducing emissions.[9] The argument that *'dirty coal'* should be phased out is rejected as its future is seen in 'low-carbon' or even 'zero-emission' gas and coal plants. The depletion of fossil resources is acknowledged but 'peak-oil' will occur and become relevant only in some decades. Today's major industrial players in fossil and nuclear energy industries are able and will '*clean up their act'* and do their bit. They are already contributing to technological innovations, secure the energy supply and continuously lower GHG emissions. Transnational oil and energy companies are the driving force in bringing down emissions as they invest billions in clean or renewable energy solutions. In that sense they are not only causing environmental problems but they are also contributing to the solution.[10] The 'alarmism' of some environmental interests is rejected. All former gloomy predictions and that the world will run out of oil turned out to be wrong. Improvements in exploration and extraction technologies have always outweighed increases in demand. Gas, coal and nuclear are well-established technologies, which have provided the consumer with secure and cheap energy for decades and they will continue to do so.

The phasing-out of nuclear energy, coal and gas production will not only lead to higher import dependency but also yield a loss in employment and deprives us of an important field of research and international competitiveness. Western countries cannot afford to lose the accumulated scientific and engineering expertise and know-how in mining, manufacturing of machines, turbines, and conventional energy technologies in international markets to strategic competitors like India or China.

Similarly the role prescribed to nuclear energy is a positive one. Nuclear power is seen as safe, clean, and cheap source of energy. The safety records and standards in Western countries are the highest in the world. It is argued that we will simply not be able to meet emission reduction targets unless nuclear energy has a part to play in a sustainable energy future. We cannot afford *not* to use nuclear power as a transition facilitator.[11]

> Every authoritative energy analysis points to an inescapable imperative: we cannot conceivably achieve a global clean-energy revolution without a rapid expansion of nuclear power to generate electricity.[12]

The future role of renewable sources of energy is portrayed as growing in relevance but limited in their overall impact. EOP is pleading for a 'realistic' evaluation of renewable energies, which are thought to be environmentally benign, but immature and inadequate in solving the energy crisis.[13] This critique rests on three pillars:

---

[9] Brian Wilson, former energy minister, currently the prime Minister's special Representative on Overseas Trade, The Guardian, 19 September 2004.

[10] For example, Lorde Browne stated that BP "has an interest in removing the sense of fear associated with environmental risks. In other words, it is prudent to be seen as part of the solution to global warming, rather than as the major contributor to its causes" (*The Guardian*, 15 September 2004).

[11] For this see for example Rothstein (1995: 278/9). Or as Gert Maichel, a Chief Executive of RWE Power has put it: "CO2 reduction as well as phasing out nuclear . . . doesn't go together" (Energy bosses urge EU to back nuclear power, www.planetark.com, 26 November 2004).

[12] See Ritch (2005). Note that he avoids the term 'sustainable' and talks of a 'clean energy' revolution. Other expressions of the nuclear EOP rhetoric are 'emission-free' or 'emission-neutral' energy systems.

[13] For example in Germany out of 121.000 MW of installed capacity 40.000 MW will have to be replaced by 2020. See Dena (2005: 3). John Loughead, Executive Director of the UK Research Energy

Beyond general doubts on the potentials of renewable energy technologies an economical and a technological line of counter argument is brought forward:
- In *general*, renewable sources of energy (such as wind, solar or biomass) do neither have the potential to meet a growing total global energy demand, nor can they compensate for the 'energy gap' that is projected to occur if nuclear and coal energy is phased-out. Renewables cannot substitute for the conventional energy sources like nuclear, oil or gas.[14]
- *Economically* renewables are too expensive and their lack of economic efficiency limits their role in the near and mid-term future.[15] Only illegitimate market-distorting state intervention by subsidies keeps them alive and enables them to expand at the expense of the consumer and taxpayer. Renewables will remain uncompetitive within the next decades and as long as fossil energy sources are still available in large quantities and at cheap rates renewables are no viable alternative to fossil and nuclear energy. Proponents of a fossil-free energy system are accused of neglecting basic laws of economics.
- *Technologically* renewables as intermittent energy sources are portrayed as incompatible with existing infrastructures (electricity grids) and the structure of energy demand (no base-load capacity) and as such represent a major challenge to grid stability and security of supply. The large-scale introduction and integration of basically decentralised renewable energy sources (for example large off-shore wind farms) raises system costs for grid modernization and costly 'back-up capacities.'[16]

The Discontinuity Story Line

Fundamentally opposed to the outlined EOP belief system is the *Discontinuity* narrative that rejects most of EOP's basic assumptions, positions and conclusions. Its response to climate change is a more active and anticipatory one (precautionary principle). Implicitly Discontinuity emphasises de-carbonization as necessary but not sufficient condition to fulfil the criteria of a sustainable energy system. As the term already suggests the Discontinuity approach is the most disrupting policy response in energy policy in that it promotes the radical substitution of all types of fossil and nuclear energy with renewables (see Scheer 1999, 2005). Discontinuity aims at the *ecological transformation* of the entire society and economy in contrast to *technological adaptation* in the EOP sense. It can be understood not only as a gradual shift in strategic alignment, but also as a fundamental interruption to the conventional energy system.

Centre states: "There is nothing wrong with renewable technology as such but if we look at the rate of penetration it is looking quite likely that it will not meet all our future energy requirements". (Do we need nuclear? *BBC News*, 26 April 2005).

[14] "It is completely utopian to assume that renewables could fill the gap left over by phase-out of fossils and nuclear. . . Renewables were to grow by the factor six, an unrealistic project". Carl Christian Von Weizsäcker described renewable energy support even as irrational and unaffordable "Eurocentric affluence hobby" (von Weizsäcker 2004: 13).

[15] Jesse Ausubel, Director of the human environment program at New York's Rockefeller University, calls renewable energy sources "false gods - attractive but powerless". They are viewed as capital and land-intensive, but "despite all the hype, tax breaks, and incentives not yet remotely cost competitive".

[16] Intermittency as the single largest driver of system costs has been identified by the ILEX study that found that "in all scenarios we investigated, we found that extending renewable generation . . . by 2020 would increase system costs". See ILEX (2002: iii). The issue that nuclear power plants rfom the beginning of their existence also needed some kind of back-up to cover sudden shut-downs is an absolute non-issue in the debate on energy futures. Instead discussions about nuclear usually are about climate change targets and economic efficiency.

Climate change in this view is a fact that can already be observed in rising global average temperatures, melting ice caps, rising sea levels and the increase of extreme weather conditions like floods and storms and draughts. There is urgent need for action to prevent the fragile global ecosystems from collapse. The 'age of oil' and cheap energy is over. Economies cannot wait until the last gallon of oil is burned. Unlike in the EOP discourse, peak-oil is not a theoretical matter that becomes relevant in a few decades: it will arrive sooner rather than later or it has even already occurred. Furthermore governments, international bodies like the IEA or OPEC are accused of misleading the public and policy-makers for decades by irrational accounts of fossil energy reserves and systematically underestimating the potential of renewables.[17] The downscaling of proven reserves by multinationals like Shell in 2004 is brought forward as evidence.

> Most official figures are grossly unreliable . . . companies seldom report their true findings for commercial reasons, and governments - which own 90% of the reserves - often lie . . . the two most widely used estimates of world oil reserves, drawn up by the Oil and Gas Journal and the BP Statistical Review, both rely on reserve estimates provided to them by governments and industry and do not question their accuracy.[18]

Discontinuity approaches the question of future energy supply from the exact opposite end than the EOP side. There is no energy gap that has to be filled in the future. Renewables not only theoretically have the potential to replace all conventional energies. The opposite it is argued is the case: If only a small percentage of the enormous untapped renewable potential can be harvested there is no energy shortage. The usage of this overabundance is just a matter of political determination and technical re-arrangement of energy infrastructures. But the most serious attack is launched at the other narratives in that the underlying energy economics are evaluated as clearly biased towards oligopolistic and quasi-competitive conventional energy sectors. The Discontinuity story line is arguing that the basic assumptions and conclusions underlying energy policy and energy economics are fundamentally flawed.

- The current energy system is mainly one that is characterised by serious policy and market failures. Energy prices fail to internalise the environmental externalities, i.e. health and other social costs.[19]
- The monopolistic energy markets are not functional and self-regulating mechanisms have to be complemented by strong regulation. Government intervention is needed and justified to correct these market failures, as only 'hard' regulatory instruments and fiscal incentives can attain the internalization.[20] On the instrument side discontinuity favours a re-regulation (as opposed to deregulation

[17] Die internationale Vebreitung Erneuerbarer Energien ist nicht mit fossilen Strukturen zu erreichen, press release *EUROSOLAR*, www.eurosolar.org, 28 November 2005.
[18] Colin Campbell, The end of oil is closer than you think, *The Guardian*, April 21, 2005.
[19] Social costs include direct costs in public health, and impacts of unsustainable energy use and more indirect effects of climate change, for example premature deaths by pollutants, chronic respiratory diseases, fatal mining accidents, or costs caused by natural disasters.
[20] Other instruments mentioned here include the penalisation of carbon emissions by fossil fuel or kerosene taxation, congestion charges, abolition of subsidies and depreciation options, more ambitious targets for renewable electricity and tax incentives for renewables, the streamlining of institutions, support of micro power technologies, R&D for grid development and enforcement and the renewable agenda should be expanded beyond electricity into the heat and transport sectors. See Manifesto by the UK renewable trade associations (2005: 2-3) and GTZ (2004) for the outcome of the international Renewables 2004 conference.

or self-regulation) of energy markets with a clear prioritization of renewable energy technologies. It is argued that an energy transition will have to be initiated by political forces because the market itself in the pursuit of short-term profit interests is blind to long-term benefits. Market solutions will not suffice to solve the energy question as long as the market itself is dysfunctional.

- Despite opposite rhetoric by major market players massive direct and indirect subsidies to conventional energies distort the energy pricing mechanism and prevent a 'level-playing field' for alternative energies. Subsidies for renewables are low compared to public support that was granted for decades for fossil and nuclear energy. 'Small is beautiful, but big is subsidised'. The alleged economic inefficiency of renewables therefore is not an economic law but result of decades of irrational funding and political support by direct or indirect subsidies. Energy has to be paid for anyway, so a strategy 'away from oil' should be initiated by encouraging investments in renewables.
- Proponents of conventional energies take the past and status quo of energy provision as basis for future energy scenarios, only look at internal system costs and thereby neglecting the major advantage of renewables: no fuel costs and hardly any environmental externalities.
- Renewable sources of energy are the only ones that display a constant reduction of costs and have the highest growth rates. High upfront costs will be compensated in the long-term. Renewable energy lobbies and business associations hold the position that "the costs of avoiding the worst effects of climate change by reducing greenhouse gas emissions today will be far lower than the cost of tackling the consequences of climate change tomorrow".[21]

Economic instruments as promoted in the Kyoto Protocol are seen as insufficient and treating symptoms instead of going to the root of the problem. The flexible mechanisms are just seen as techno-fix that merely manages emissions but avoids or delays the hard decisions on crucial system changes. In fact even the moderate targets of reducing emissions by six percent in industrialised countries are often missed.[22] The same is true for energy efficiency measures, which are important but not going far enough. In both cases the unsustainable fossil and nuclear resource basis is left untouched.

> The focus on efficiency gains is often seen as wildly optimistic when all current experience suggests that, in most areas, efficiency gains per unit of consumption are usually outstripped by overall increases in consumption (Barry/Paterson 2004: 770).

The dissemination of renewable energies cannot be achieved within large-scale and centralised fossil and nuclear infrastructures. The transformation to a 'solar economy' itself cannot be initiated by the same players that most benefit from the status quo in the current energy system. Today's oil and energy industries will and cannot be the driving force behind processes of fundamental change. A transition is delayed, postponed by the 'nuclear and fossilist establishment' (H. Scheer). Letting the corporate energy industries set the pace of a transition is seen as a danger, because

[21] Stephen Cox, Executive Secretary of the Royal Society (2005: 34).

[22] In 2004 global CO_2 emissions accounted for 27.5 billion tonnes, that is 4.5% more than compared to 1990 levels. In OECD countries emissions increased by 16%, in the EU-15 countries by 4.5%, and China increased its emissions in 2004 by 15% compared to 2003.

their core businesses are still in the conventional sector. Image campaigns and advertisements cannot hide the fact that expenditure on research and development in renewables are 'quantités négligables' compared to investments in new exploration sites and conventional infrastructures and just sideshows for greening the corporate image.

Discontinuity Perceptions of Fossil Fuels and Nuclear Energy

Fossil fuels are identified as origin and not the solution of the energy crisis. Discontinuity in its most radical form argues that there is no such thing as clean fossil energy or save nuclear energy. From a Discontinuity perspective new large-scale technologies are questionable 'solutions'. One the one hand they are inefficient, as they do not correspond with the decentralised nature of energy consumption. On the other hand environmental groups fear that new conventional technologies (like fusion reactors or carbon capture and sequestration) will lock-in old-fashioned technology and thereby preventing investments in new energy sources.[23] Beyond that these technologies are associated with uncertain environmental and economic risks. '*Clean coal'* is an oxymoron that is used to keep the coal industry's dirty business alive.[24] Natural gas, that is often portrayed as clean solution usually has to be imported from the same countries and than oil and does thereby not reduce import dependencies. As fossil energy source it is also subject to depletion and if leaked to the atmosphere an even more potent greenhouse gas than CO_2.

Nuclear energy never was and never will be economically viable without government support. Therefore the often-claimed 'competitiveness' of nuclear power is challenged, and alongside with unsolved problems of waste management and risks associated with nuclear energy (accidents, proliferation, low-level radiation in everyday operation, terrorism and social acceptance) disqualifies nuclear energy as being sustainable in any way. When evaluated in a total life cycle analysis it is not even considered to be carbon-free. The fossil and nuclear energy path has reached a dead end and only renewables such as solar, wind, biomass, hydro, geothermal, tidal, wave energy) can be seen as truly sustainable sources of energy. Along with the de-centralization of the energy system goes the creation of virtually millions of new jobs, as renewables are more labour intensive in all parts of the energy chain (see for example Pfaffenberger et al. 2003). The loss of jobs in other sector is seen as minor problem compared to the benefits a renewable energy system would generate: the revitalization of the agricultural sector, consumer empowerment, and re-democratization of energy supply. The transition is not only a matter of future generations but can be achieved within one or two decades if there is a strong political commitment.

The Continuity Narrative

Finally, the Continuity discourse is situated between the EOP and the Discontinuity approach and as such holds positions that often seem to be a pragmatic compromise between the more extreme policy responses. It integrates elements of both, maintaining conventional structures as well as moderate system changes. The Continuity position can be best described as supporting a resource basis that remains mainly fossil and nuclear, although the support for renewables is somewhat more

[23] Erdgasspeicher bei Berllin, sandstein soll CO2-Grab warden, www.Spiegel.de, 20 November 2005.
[24] Greenpeace, Clean coal not possible, www.sunstar.com, 29 January 2005.

pronounced than in the EOP case in that it incorporates elements and guiding policy principles of *ecological modernisation*. Characteristic for a Continuity position is that "it is now almost universally accepted . . . that problems such as climate change require fundamental, long- term solutions. . . Critically, the solutions rely less on end-of-pipe' technical measures such as pollution abatement and more on transforming the whole of society and the economy" (Jordan 2000: 259).

The Continuity belief system acknowledges that climate change is a fact and a threat to human development. Continuity promotes the idea that economic growth, social dimensions, and environmental concerns like air pollution and climate protection can be reconciled and do not exclude each other. The main drivers of a transition are climate change and the international commitments of CO_2 reductions. The nature of a transition is understood as gradual market and technology-induced innovation shifts towards the ultimate aim of a de-carbonised economy or that of a 'low-carbon energy system'.[25] The most distinct feature of the continuity style of thinking though, is the focus on energy efficiency, sometimes called the 'forgotten pillar' of energy policy. The easiest and way of reducing emissions it is claimed is saving energy, increasing energy efficiency or the 'rational use of energy'.[26] The iconic representation of this is the 'Factor 4' approach that aims at halving energy use and simultaneously doubling energy intensity. The ultimate and cheapest way of emission reduction is not using energy at all or get the most out of it, and thereby diminishing the *'ecological footprint'* of individuals and economies.

Inherent to the Continuity style of thinking is a strong belief in the self-regulating functions of market forces and market instruments in energy policy. On the instrument side 'market-based' or 'market-near' solutions are clearly preferred to state interventions. The Kyoto Protocol and emission-trading schemes are setting the right incentives in the fight against climate change yielding substantial emission reductions in the most cost-effective way. From a Continuity perspective it would be fully acceptable that centralised plants using domestic coal (for example for hydrogen production) release carbon dioxide as long as it is in accord with purchased future emission credits (Cherry 2004: 126). Although Government has the key role to play by providing consistent market signals, state intervention is only acceptable in terms of securing energy supply or risk mitigation. Energy policy only has to provide stability and predictability of energy investments and a framework that is filled by the interaction of competitive market players. The main concept therefore is not the preferential treatment of a certain technological innovation or a complete system change but a gradual modification of the energy chain. Self-regulation of industrial players (for example the automotive industries' commitment to reduce the average vehicle fuel consumption) are welcome initiatives.

[25] Note that the wording of 'de-carbonizing' or 'low-carbon' contains notions of conventional energy sources in contrast to the ideal of the Discontinuity approach that wants a complete system change to a 'non-carbon' or a 'solar-economy'.

[26] In a discussion paper the Wuppertal Institute (2004) defines a sustainable energy future as 'decentralised, linked and low-risk'. Examples for the Factor 4 approach include better insulation in households, the three litre car, awareness campaigns. Other Continuity measures and technologies include improved efficiency in energy conversion, combined heat and power, fuel cells, substitution of burners by gas turbines, heat pumps, or sterling engines.

Typically Continuity supports add-up technologies or other energy-efficient 'bridging technologies' that fit into existing energy infrastructures but also pave the way for future innovations. 'Alternative' transport fuels (bio diesel, sun diesel, ethanol) or gas-electro-hybrid cars in the transport sector are illustrations for the Continuity style of thinking as they combine existing technologies with potentially more disruptive innovations. Another example is hydrogen technology.

> Because hydrogen can be made from many different sources, a future hydrogen energy system could evolve in a variety of ways. In industrialised countries, hydrogen might get started by 'piggybacking' on the existing energy infrastructure (Ogden 1999: 63).

Fossil, Nuclear, and Renewable Energy in the Continuity Perspective

Concerning the future of energy sources all options, including nuclear fission and fusion energy should be 'kept open'. In terms of energy sources Continuity prefers the whole spectrum of available and yet to be developed energy sources and carriers.[27] A sustainable energy system is one that relies on a 'broad energy mix', a variety of options including conventional and renewable sources, centralised and decentralised infrastructures and expectations for new energy forms such as bio fuels or the fusion reactor. A common place of the Continuity narrative is that 'there is no magic bullet'. Consistent with a Continuity belief system is the notion that an intelligent policy and instrument mix in energy policy consists of the usage of all thinkable leverages and solutions including diversification of primary energy sources and technologies, mobilization of energy saving potentials and energy efficiency, R&D expansion to secure energy security (DB Research 2004: 22). Continuity offers a more positive account of renewables. In principle renewable sources of energy offer the best and most sustainable solution to the energy problem, with minimal air pollution, but all energy futures have to be competitive and cost efficient compared to conventional or nuclear fuels. So far renewables do not fulfil these criteria yet. Attitudes reflecting Continuity elements include that energy demand reduction measures will be as important as generation technologies, and will require both technological and behavioural changes. Beyond that energy efficiency and renewables are two sides of the same medal, a tandem that guarantees maximum efficiency with least cost options and fewer subsidies. Renewables according to the Continuity interpretation can only be successful if energy consumption is reduced and energy efficiency measures have been successfully implemented on a large scale. In that respect there is a closer link to the Discontinuity section that not only wants to bring down fossil fuel use but increase the share of renewables as fast as possible in contrast to the EOP school that focussed on the option of technological emission abatement in improved conventional energy sources.

Continuity's position on nuclear power is shaped by economic rationalism. It is neither seen as inevitable (as in the EOP belief system), nor as an option that has to be prevented by any means (as the Discontinuity fraction demands). If nuclear power demonstrates economic and ecologic efficiency it is a technology worth considering and supporting. Nuclear power is thus seen as, lesser evil in the context of climate

[27] Such a position is taken over by the European Commissioner Potocnik. "I do not believe we can pin all our hopes on one single solution. It is our responsibility to keep options open, for ourselves and for our children and grandchildren. We should strive to have a sustainable energy mix. In order to achieve this, we need a comprehensive research effort looking at a broad range of energy technologies: from renewables, through clean coal, to nuclear fission and fusion" (Janez Potocnik, European Commissioner for Science and Research, Nuclear Energy for a New Europe 2005, 5 September 2005).

change. Nuclear can be an important source of carbon-free electricity in the future but the economics (nuclear waste disposal and decommissioning of facilities) and risks have to be carefully considered.

Summary and Conclusion

Three narrative strategies have been presented and distinguished. EOP pursues the track of economies of scale, and highlights the use of improved well-established technologies. Continuity promotes the rational use of energy, a diversified energy mix. Until the long-term target of a renewable energy system is achieved, a combination of large and small-scale technologies, bridging technologies is favoured within a market based policy approach. Discontinuity wants a paradigm shift in energy policy towards decentralised and small-scale renewables. Market forces alone cannot achieve that. Political support and commitment and clear priorities are needed.
All outlined belief systems have their strengths and weaknesses. The EOP approach has to be characterised as the one that represents the largest lock-in potential. It is usually presented by actors that are linked or close to fossil or nuclear energy companies and policy. It is a rather reactive response to the challenges ahead. On the other hand the EOP approach guarantees stability and prevents sudden interruptions that are unwanted in energy economics and long-term planning. More still than Discontinuity the End-of-Pipe approach is firmly based on decades of scientific and engineer knowledge and expertise and less influenced by ideological terminology. On the other hand it can be seriously questioned if the targeted emission reductions can be achieved by business-as-usual policies?

The Continuity story line is integrative and politically the most pragmatic approach. The narrative's assumptions can be seen to be closest to the ideal of a balanced interpretation of sustainable development that incorporates social, economical, and ecological aspects. The attractiveness of the Continuity system thus stems from the advantage that this paradigm provides the best options for pragmatic solutions that can be politically implemented without too many conflicts. Its vagueness can be filled and interpreted even by mutually exclusive energy and sustainability assumptions. No interests are really touched, all is integrated into an 'overall framework' where nuclear energy, renewables, old and new industries and technologies can peacefully co-exist and find their niches. But the demand for cooperation and increasing energy efficiency indirectly supports tendencies towards conserving the status quo, at least obstructs more sudden shifts in the energy mix of a society. And a Continuity approach that uncritically attempts to integrate the various dimensions will soon be confronted with serious target conflicts (affordable energy costs, competitiveness, economic growth and ecological integrity, fuel diversity, employment, energy independence, environmental goals and international commitments) that cannot all be fully attained at the same time. Possibly Continuity underestimates the challenges ahead as the target conflicts are not as easily reconcilable as the narrative suggests.

The Discontinuity belief system presents the most idealistic and most disruptive response. Whereas the EOP paradigm overestimates the economics, the Discontinuity fraction goes too far in its positive evaluation of a complete energy transition, neglecting social or economic side effects. It is also not consistent to radically criticise state intervention and at the same time ask for 'commitment' and 'stable incentives structures', which, in the end, are also forms of direct or indirect subsidies. Yet it presents valid critiques on current energy system configurations, but it might be

possibly too ambitious in scope, targets, and timeframes. Too much is wanted in a too narrow time scale. A solar economy in the end will also have to rely on economic reason. Despite possibly being too optimistic the Discontinuity perspective points at some serious shortcomings. Altogether the Discontinuity narrative is the environmentally most benign and most desirable outcome and 'ecological effectiveness' cannot be achieved by politics of 'muddling through'. It is the most aggressive response and quite realistic in its assessment that any kind of transition will only occur after hard battles and political conflicts. If climate change and GHG emissions are seen as the biggest environmental problem of humanity, more than some technical modifications and clean energy rhetoric will be needed to avoid another global economic crisis caused by energy shortages.

Notes

Some literal quotes have been translated from German by the author.

References

Barry, J. and Paterson., M, (2004) Globalisation, Ecological Modernisation, and New Labour, *Political Studies*, 2004, vol. 52, pp. 767-784.

Bartelmuss, P (1999): *Sustainable Development- Paradigm or Paranoia*? Wuppertal Paper Nr. 93, May 1999, (www.wupperinst.org).

Bechberger, M. and Reiche D., (2006) *Ökologische Transformation der Energiewirtschaft, Erfolgsbedingungen und Restriktionen*, Erich Schmidt Verlag: Berlin.

Berkhout, F., (2004) Technological regimes, path dependency and the environment, *Global Environmental Change* 12(1): pp1-4.

Brand, K. and Jochum, G., (2000) Der deutsche diskurs zu nachhaltiger Entwicklung, MPS-Texte 1/2000, Abschlussbericht eines DFG-Projekts zum Thema "Sustainable Development- zur sozialen Konstruktion globaler Handlungskonzepte im Umweltdiskurs, Münchner Projektgruppe für Sozialforschung e.V., München.

Casten, T. and Downes, B., (2005) Critical Thinking about Energy. The Case for Decentralised Generation of Electricity, *Skeptical Inquirer*, January/February 2005, (www.csicop.org).

Cherry, R S., (2004) A hydrogen utopia?, *International Journal of Hydrogen Energy*, vol. 29, pp. 125-129.

DB Research (2004) Aktuelle *Themen, Energieperspektiven nach dem Ölzeitalter*, 2 December 2004, (www.dbresearch.de).

Dena (2005) Energiewirtschaftliche Planung für die Netzintegration von Windenergie in Deutschland an Land und Offshore bis zum Jahr 2020 (dena-Netzstudie). Zusammenfassung der wesentlichen Ergebnisse der Studie durch die Projektsteuerungsgruppe, Deutsche Energie Agentur (dena), 23 February 2005, (www.deutsche-energie-agentur.de).

DPG (2005) Klimaschutz und Energieversorgung in Deutschland 1990 – 2020, Eine Studie der Deutschen Physikalischen Gesellschaft (DPG), Bad Honnef, September 2005, (www.dpg.de).

DTI (2003) *Energy White Paper. Our energy future - creating a low carbon economy*. Presented to Parliament by the Secretary of State for Trade and Industry by Command of her Majesty, Norwich, February 2003, (www.dti.gov.uk/energy/whitepaper/index.shtml).

EWI (2005) Ökonomische Auswirkungen alternativer Laufzeiten von Kernkraftwerken in Deutschland. Gutachten des energiewirtschaftlichen Instituts an

der Universität zu Köln (EWI) und der Energy Environment Forecast Analysis GmbH, Berlin, (www.ewi.uni-koeln.de/content/forschung_und_beratung/projekte/index_ger.html).
Feindt, P H. and Oels, A., (2005) Does Discourse matter? Discourse Analysis in Environmental Policy Making, *Journal of Environmental Policy and Planning*, vol. 7, no. 3, pp. 161-173.
Fischer, F. and Forester, J., (eds.) (1999) *The Argumentative Turn in Policy Analysis and Planning*, London: University College Press.
Foray, D. and Grübler, A., (1996) Technology and the environment: An overview,*Technological Forecasting and Social Change* vol. 53, no. 1, pp. 3-13.
Grübler, A., Nakicenovic, N and Victor, D G., (1999) Dynamics of energy technologies and global change, *Energy Policy* vol. 27, pp. 247-280.
Geels, F. (2002) Towards socio-technical scenarios and reflexive anticipation: Using patterns and regularities in technology dynamics. *Shaping technology, guiding policy: concepts, spaces and tools*, K. H. Sorensen and R. Williams, Cheltenham, UK: Edward Elgar. pp. 255-281.
Geels, F. W. and Smit, W. A., (2000) Failed technology futures: pitfalls and lessons from a historical Survey, *Futures* vol. 32, nos. 9-10, pp. 867- 885.
Goltsov, V. A. and Veziroglu, T. N., (2001) From hydrogen economy to hydrogen civilization, *International Journal of Hydrogen Energy*, vol. 26, no. 9, pp. 909-915.
Gosselink, J. W., (2002) Pathways to a more sustainable production of energy: sustainable hydrogen—a research objective for Shell, *International Journal of Hydrogen Energy*, vol. 27, nos. 11-12 , November-December 2002, pp. 1125-1129.
GTZ (2004): Conference report. Outcomes & Documentation – Political Declaration/ International Action Programme/ Policy Recommendations for Renewable Energies. Renewables Conference Secretariat, Deutsche Gesellschaft für technische Zusammenarbeit (GTZ) (ed.), August 2004, (www.renewables2004.de/pdf/conference_report.pdf).
Hajer, M. A., (2000) *The Politics of Environmental Discourse. Ecological Modernization and the Policy Process*. Oxford: Clarendon Press.
Hajer, M. and Veersteeg, W., (2005) A decade of Discourse Analysis of Environmental Politics: Achievements, Challenges, Perspectives, *Journal of Environmental Policy and Planning*, vol. 7, no. 3, pp. 175-184.
Héritier, Adrienne (ed.) (1993): Einleitung. Policy Analyse. Elemente der Kritik und Perspektiven der Neuorientierung. Policy Analyse. Kritik und Neuorientierung, PVS Sonderheft 24, Opladen, pp. 9-36.
ILEX (2002) Quantifying the system costs of additional renewables in 2020, a report to the DTI, ILEX energy consulting, in association with Professor Goran Strbac, Manchester Centre for Electrical energy.
Jacobsson, S. and Laube, V., (2006) The politics and policy of energy system transformation — explaining the German diffusion of renewable energy technology, *Energy Policy*, volume 34, no. 3 , February 2006, pp. 256-276.
Jänicke, M., Kunig, P and Stitzel, M., (1999) Lern- und Arbeitsbuch Umweltpolitik. Politik, Recht und Management des Umweltschutzes in Staat und Unternehmen. Dietz: Bonn.
Jordan, A., (2000) Environmental Policy, in Dunleavy, P., Gamble, A., Holliday, I. and Peele, G (eds.)., *Developments in British Politics 6,* Basingstoke: Palgrave.
Legett, J., (2001) *The Carbon War. Global warming and the end of the oil era*. New York: Routledge.

Lovins, A., Kyle D., Odd-Even., B., Koomey, J. and Glasgow, N J., (2005) *Winning the oil endgame: Innovations for profits, Jobs, and security*, Rocky Mountains Institute, Snowmass, Colorado, (www.rmi.org**,** www.oilendgame.com).
MacRae, D., (1993) Guidelines for policy discourse: Consensual versus Adversarial, in Fischer and Forester (eds.), pp. 291-318.
Martin, J-M (1996) Energy Technologies: Systemic Aspects, Technological Trajectories, and Institutional Frameworks, *Technological Forecasting and Social Change* vol. 53, pp. 81-95.
Matthes, F. C. and James, M., (2002) Energy Turnaround 2020 -The path to a sustainable energy system in Germany, An Öko-Institut Study, World Summit Papers of the Heinrich Böll Foundation, No 15, published by Heinrich Böll Foundation, Berlin.
Pfaffenberger, W., Nguyen, K. and Gabriel, J., (2003) Ermittlung der Arbeitsplätze und Beschäftigungswirkungen im Bereich Erneuerbarer Energien, Bremer Energie Institut (BEI), Bremen, Dezember 2003. (www.bei.de).
Ritch, J., (2005) The Key to Our Energy Future, Washington Post, 26 April 2005, (www.washingtonpost.com).
Reichel, W., (2004) Kohle bleibt unverzichtbar. Der Ausstieg aus der Steinkohleförderung wäre ein Fehler, *Internationale Politk*, vol. 8, pp. 49-52.
Roberts, P., (2005☺ *The End of Oil: The decline of the petroleum economy and the rise of a new energy order*, London: Bloomsbury.
Roe, E., (1994) *Narrative Policy Analysis*, London: Duke University Press.
Rothstein, J., (1995) Nuclear Energy, Environmental Problems and the Hydrogen Energy Economy, *International Journal of Hydrogen Energy*, vol. 20, no. 4, pp. 275-281.
Royal Society, (2005) Royal Society Response to DEFRA Review of the UK Climate Change Programme, Policy Document 02/05, April 2005.
Sabatier, P. A., (1993): Advocacy-Koalitionen, Policy-Wandel und Policy-Lernen: Eine Alternative zur Phasenheuristik, Heritier, Adrienne (Hg.) Policy Analyse. Kritik und Neuorientierung. PVS-Sonderheft 24, Opladen, pp. 116-148.
Sabatier, P A. and Jenkins-Smith, H. C., (1999) The Advocacy Coalition Framework. In: Sabatier, Paul A. (ed.), *Theories of the Policy Process*. Boulder, Colorado: Westview Press. pp. 117-166.
Scheer, H (1999): Solare Weltwirtschaft. Strategie für die ökologische Moderne. Kunstmann, München.
Scheer, H (2005) Energieautonomie, Eine neue Poltitik für erneuerbare Energien,Kunstmann, München.
Schwartz, P. and Reiss, S., (2005) Nuclear Now! How clean, green atomic energy can stop global warming, *Wired Magazine*, 13 February 2005 (www.wired.com/wired/archive).
UK Renewable energy trade associations (2005) The Renewable Energy Manifesto (www.britishhydro.org/Renewable%20Energy%20Manifesto%20(Feb%2005).pdf).
Unruh, G. C., (2000) Understanding carbon lock-in, *Energy Policy* vol. 28, pp. 817-830.
Unruh, G. C., (2002) Escaping carbon lock-in, *Energy Policy* vol. 30, pp. 317-325.
von Weizsäcker, E U., Lovins, A B. and Lovins, H L., (1996) Faktor Vier,Doppelter Wohlstand-halbierter Naturverbrauch. Der neue Bericht an den Club of Rome, Knaur, München.
von Weizsäcker, C., (2004) Der teure Heiligenschein erneuerbarer Energien, *Internationale Politik*, August 8/2004, pp. 7-16.

WBGU (2003) Welt im Wandel. Energiewende zur Nachhaltigkeit. Wissenschaftlicher Beirat der Bundesregierung ‚Globale Umweltveränderungen' (ed.), Berlin,(www.wbgu.de/wbgu_jg2002.pdf).
WHO (2002) *The World Health Report 2002, Reducing Risks, Promoting Healthy Life*, Geneva, (www.who.org).
Wuppertal Institute (2004): Wege von der nachholenden zur nachhaltigen Entwicklung. Infrastrukturen und deren Transfer im Zeitalter der Globalisierung. Welche Globalisierung ist zukunftsfähig?, Wuppertal paper 140, February 2004, (www.wupperinst.org).
Wuppertal Institute (2005): Ökologische Einordnung und strukturell-ökonomischer Vergleich regenerativer Energietechnologien mit anderen Optionen zum Klimaschutz, speziell der Rückhaltung und Speicherung von Kohlendioxid bei der Nutzung fossiler Primaerenergien. 1. Zwischenbericht, Wuppertal Institut fur Klima, Umwelt und Energie GmbH, Wuppertal, Stuttgart, Potsdam, (www.wupperinst.org), May 2005.

Design of Cutaneous Foot-based Virtual Terrain Recognition System

M S Fipaza and D G Caldwell

Centre for Advanced Robotics and Automation

Introduction

The primary objective of VR and telepresence applications in haptics is to create an illusion of immersion and improve interactions between humans and machines. This illusion should ideally affect all five human senses, but research developments have over the years significantly advanced in providing tools that trigger visual and auditory senses. Latest developments in haptic interface technology however complement audio and visual information techniques in VR.

Haptics are subtle sensory information concerned with the sense of touch. The sensation has two complex components, including cutaneous and kinaesthetic. Cutaneous (skin) sensation is mediated by variety of nerves which respond to displacement, pressure, temperature and vibrations. Kinaesthetic/ Proprioceptive sensing is primarily associated with joints and muscles which respond to motions and forces exerted through interactions of the body with external environment. A complete haptic sensation and/ or recognition of objects require the presence of both components (Caldwell 2005). Although numerous research efforts produced workable cutaneous sensation prototypes for fingertip tactile, full body haptics (including locomotion interfaces) is a relatively unexplored area (Fipaza 2005).

An important consideration in the cutaneous system design is the selection of the simulation mechanism and the range of nerves to be triggered for an adequate feeling. The process requires understanding of the human neuro-physiological structure and interactions. In many fingertip applications however, mechano receptive nerves are simulated for pressure and vibrations based on techniques including: air or water jet displays, mechanical tacton elements (Sarakaglou, Tsagarakis and Caldwell 2005), SMA tactile array (Fisch et al.), electro tactile, thermo-simulating devices, electro-rheological fluids (Taylor, Hosseini-Sianaka and Varley 1996), and pneumatic actuated systems (Caldwell, Tsagarakis and Gesler 1999). It is reported that Moy et al proposed a specification based on SA I signal response for tactile displays and this approach formed the design basis for most systems (Sarakaglou, Tsagarakis and Caldwell 2005). However, portability, size and flexibility were some of the challenges of earlier displays. Although there has been less development of foot displays, techniques applied in previous work on simulation of fingertip tactile are relevant in this design. The primary considerations in the achievement of this system include; low cost design, good sensitivity performance and minimum weight actuation mechanism. A successful design and construction of a truly portable, low cost and light dc motor actuated tactile display (for texture and shape) was unheard of, until a recent breakthrough at the University of Salford (Sarakaglou, Tsagarakis and Caldwell 2005).

A recent study established distribution of foot mechano-receptive nerves as follows (Kennedy and Inglis 2002): 14% SA I, 15% SA II, 59% FA I and 14% FA II. The

receptors are widely distributed with predominantly isolated, varying receptive fields and no accumulation on toes. The findings suggest that skin receptors in the foot sole behave differently to those found in the glabrous skin of hand. A Japanese researcher (Masako, Michimaru and Kouchi 2003) found no significant differences between foot sensitivity of males compared to females on the basis of Semmes-Weinstein monofilament tests for mid-aged people. Recent foot systems include the 7-DOF per foot 'novel haptic walker' that simulates various terrain transversion mechanisms (Schmidt et al.), and the foot-pedal walking simulator developed at the University of Salford (Abulgasem 2004).

The design presented in this paper is an effort to address issues in cutaneous sensation on the foot sole.

Mechanical Design

The focus of this design is to develop a system with good sensory performance levels in a display that is foot wearable. Our approach is to build an arrangement of pulsating pins that display terrain (shape, hardness) information to sole. The pins are made of polished steel each with the following dimensions; 5mm based diameter, 12mm cylindrical length with diameter of 6mm (pin body) and 2mm tip of pin has been cone shaped to enhance sensitivity. Stronger pins are required for adequate sensation on loaded (harder) skin of the sole. Initial tests achieved a force output ranging up to 30N per tacton, which is limited by air inflow pressure and base area of pin. The original specification required pins to achieve 3mm displacement at 20mm spatial separation. This would provide adequate sensation at bandwidth frequency of 2.4Hz. Each tacton is safely positioned, relatively free to move (neglecting friction) and can just fit into pneumatic cylinder and onto an inflatable balloon. The structure allows for smaller display components attached to the shoe and achieves good force output using remotely located pneumatic actuation. Figure 1 indicates the balloon in pneumatic cylinder, and lid.

Figure 1: Balloon in cylinder , pin and lid

The overall dimension of the array (pin arrangement) indicated in Figure 3 will be 73 mm x 60 mm x20 mm. Preliminary trials suggests that it would be neater for pin casing cylinders to be embedded in shoe sole as shown in Figure 2.

a) **b)**

Figure 2: a) Top of pin-cylinder embedded shoe and b) extended shoe sole

One shoe embedded pin was tested to determine feasibility of the design concept to simulate sensitivity to varied actuation forces as shown in Figure 2. Once fully developed, the display will consist of 12 pins. The layout of the envisaged array of pins is indicated in Figure 3, but would be more foot shaped.

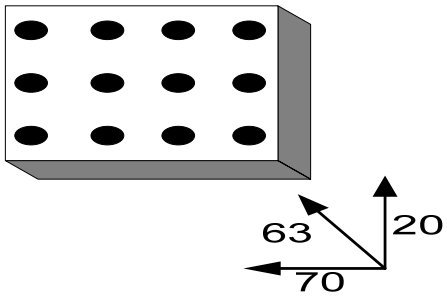


Figure 3: Array foot pin display

Pneumatic piping entry shall be on outer foot to ensure neat bundling and safe walking environment.

Actuation System

Pneumatic inflation of in a cylinder balloon serves as primary actuation method and is achieved by applying pressure to base of a pin. Each pin produces an equivalent force to foot sole subject to equivalent pressure response of the input signal. The pneumatic supply pipe to each pin balloon is made from a flexible polyethylene with inner diameter of 4mm. Pneumatic inflow to each pin balloon achieves actuation. Initial tests indicate good force output of 30N at 400 kPa. The force conforms to specification requirements and applicable measurements are discussed in the performance section. The air flow control valves included for filling and venting are low power consumption proportional valves (JOUCOMATIC 630 piezotronic series). An arrangement of the valves with loading of 960g per foot will be located remotely in the lower limb with neatly packed piping connecting to the shoe display. Each piezotronic valve has a mass load of 40 g and every pin actuation requires two independent valves.

Control System

The overall control system can be divided into the following:

i) The control PC
ii) The pressure control unit

shows the block diagram of the pressure control unit, including control board, PC control, pressure sensor and valves. Schematic pneumatic details for each pin are shown in figures 4 and 5.

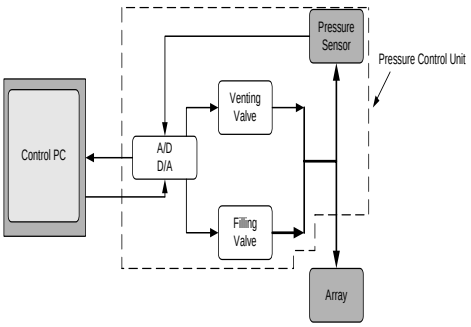


Figure 4: Diagram of pressure control unit

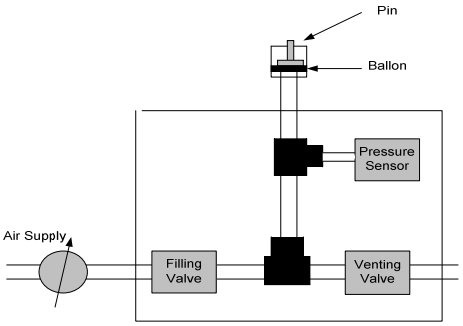


Figure 5: Schematic of pneumatic circuit of each array element

Control PC

A Pentium 4, 3 GHz PC was found to be adequate to satisfy the control requirements of the tactile pins. It is equipped with a dedicated ADC/ DAC control board to provide sensor readings from pressure sensors in the pin/ cylinder control lines and to drive the air valves controlling the elements. The microcontroller is an 8-bit Atmel processor, type ATMega 8. The RS232 serial link interfaces the control board with the PC for data communication.

Pressure/ Force Generation Unit

The unit consist of two pneumatic units including one valve for filling and another for venting. The valves have capacity to control relatively small volumes of air flow. They are unidirectional with flow rate of 6 l/min at 600 kPa (air pressure), as shown in Figure 6. A computer generated input signal controls air flow. Pin actuation force is estimated on the basis of the pressure/ force relationship:

$$P = F / A \qquad (1)$$

The pressure sensor is the 100 psi (ASDX100 series), calibrated and temperature compensated with analogue output voltage of 0.5-4.5 V.

The system responds to pressure commands based on the C++ software implementation of the PID closed loop control.

Empirical Analysis

Foot contact pressure is estimated to be the same as balloon pressure and air inflow controls actuation force. Contact pressure impact at 2-bar actuation on body mass was evaluated based on pressure, force and pin surface area relationship in (1). The 2-bar (pressure) is equivalent to 2.0387 kg/cm^2. It would therefore seem possible to lift a person weighed less than 38.4 kg at simultaneous and full actuation (2-bar) of all pins (12 pins/ foot). Results for similar calculations to determine the number of pins required to lift persons of different mass loads are as follows:

Mass (kg)	Pins/Foot
64	20
94	29

Piezotronic valves were tested for pneumatic flow at 35 V and in a range of pressure (bar) settings. The measured flows were assumed applicable to filling whilst estimating venting with consideration of atmospheric pressure. Figure 6 is a plot of flow (l/min) with respect to pressure (bar). Although pressure and flow rate have a direct proportional relationship, venting is relatively slower but is still proportional.

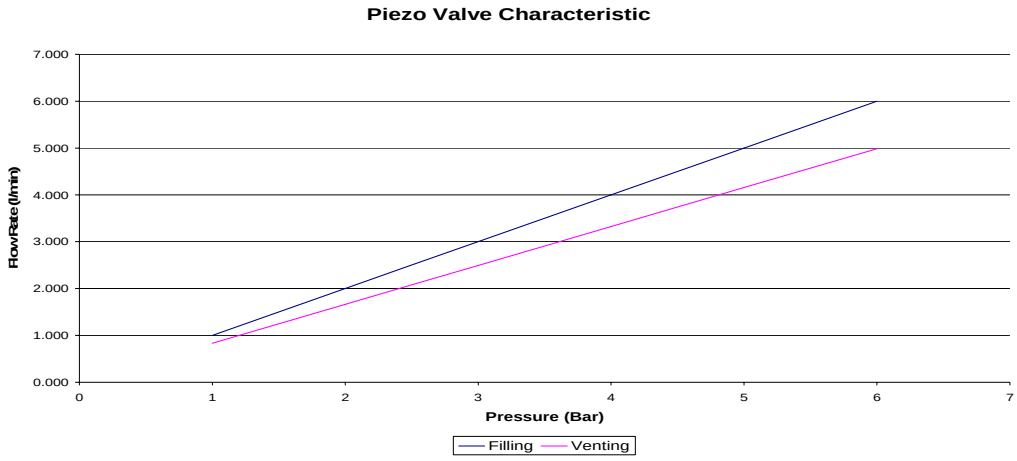


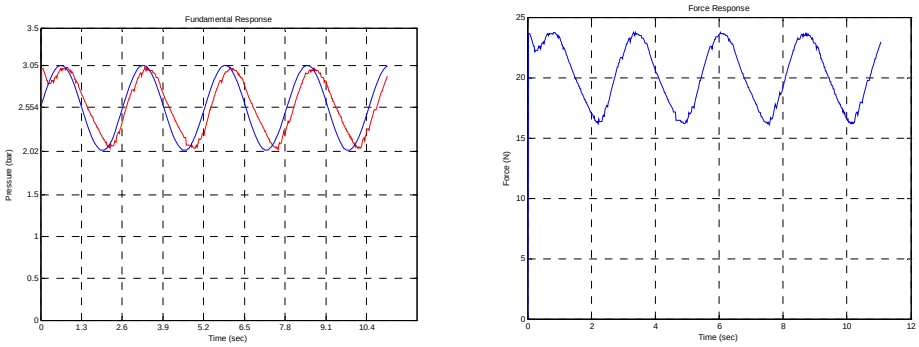
Figure 6: Estimate valve flow rate

The piezo valves are advantageous for the design because of low cost, reliability and ability to control smaller volumes of pneumatic flow as shown in Figure 6.

System Performance

Performance of the single pin system was evaluated using the MatLab/ SimuLink simulation application. Design specification allows for system performance assessment in a variety of perspectives, including:

i) *Pressure/ Force generated by each pin*: The design specification requires that a single pin has actuation force of up to 30 N for good sensation. This maximum allowable force threshold was determined by trial tests. The fundamental sinusoidal pressure input, and response signals for contact pressure of 300 kPa is in Figure 7a and the corresponding force profile with maximum of 24 N is shown in Figure 7b. The force profile is in accordance with equation (1).

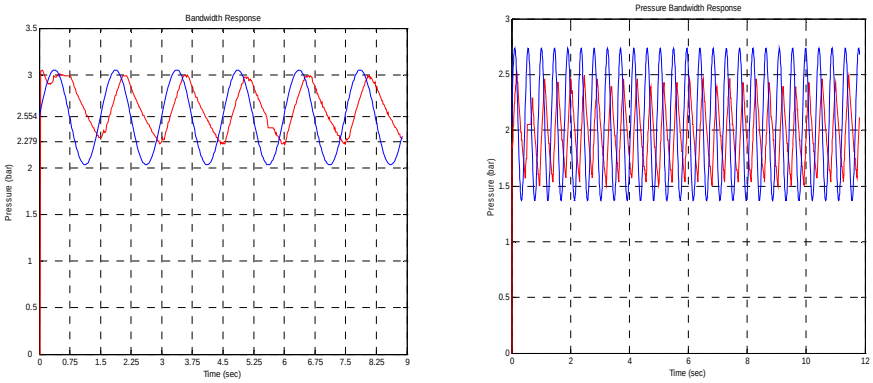


a) b)

Figure 7: a) fundamental pressure input signal and response (300kPa), b) force response signal

The response exhibits good tracking of the input signal, although there are difficulties to achieve similar profile with 0- kPa reference pressure level. There was poorer signal tracking with 0- kPa pressure reference. The attainment of 0-kPa response pressure is expected to significantly improve sensation on foot. The actuation force would exceed specification requirement above 400kPa (30N) and cause pain to user. As a result, a reasonable operating pressure limit is set at (400kPa) at this stage of the development.

ii) *Pressure Control Bandwidth*: A series of input signals at various frequencies were input to the system to test for bandwidth. Initial results achieved a low bandwidth of 0.4 Hz as shown in Figure 8a. Control software adjustments improved attainable reference pressure response and bandwidth increased to 2.4 Hz as shown in Figure 8b.



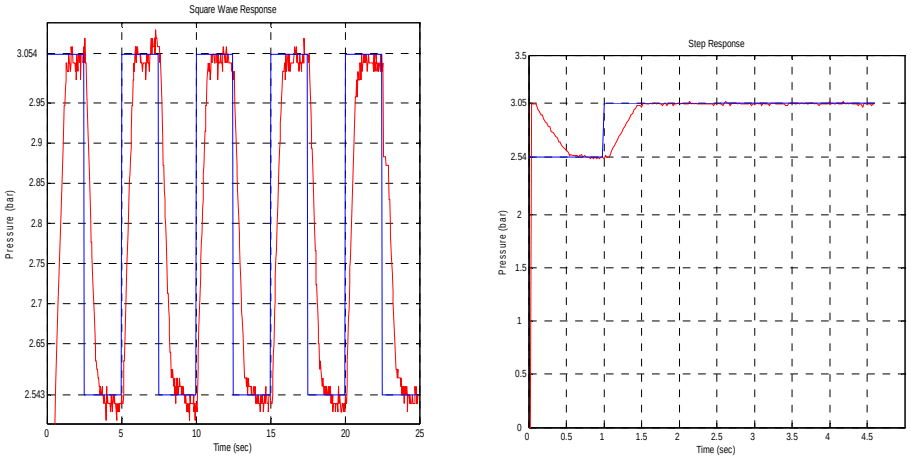
a) b)

Figure 8: a) 0.4 Hz Bandwidth, b) 2.4 Hz Bandwidth

Further developments and tuning of the control circuit are expected to increase the bandwidth for adequate simulation of all SA receptors to replicate geometrical edges and shape.

iii) Stability/ Step Response Test: Square wave and step inputs were used to test the system for stability and speed response. Test results are shown in Figure 9. A fast response with good damping was obtained for a step input, whilst good tracking with minor fluctuations at peaks for square waves was noticed. Further tuning and

adjustment of the PID controllers may address these oscillations. A sudden foot sensation with sustained force was experienced in the case of step input.



a) square wave profile b) step response

Figure 9 : Stability and Step response signals

iv) *Spatial Separation*: The original design requirement was for pin tactile system within 20mm sensitivity limits. Previous studies suggest foot receptive fields within this sensitivity threshold [7]. In this development however, this design aspect could not be verified with one pin.

v) *Displacement*: The original specification requires pin displacement of 3mm at the maximum permissible operating actuation force of 30 N. No pin displacement measurements were taken during this phase of the development as no position sensor could be included during the system operational testing. It may be sufficient to estimate displacement on force / terrain stiffness evaluation basis in future developments.

vi) *Surface Recognition:* Toolkit environment software/ virtual world to be included in future.

Conclusions

This paper outlined developments in the design of a foot based haptic system to simulate terrain edges and geometric shapes. Much emphasis on relevant previous work in tactile research concerns the fingertip. This is a relevant starting point, although knowledge of foot sensation is largely based on indirect evidence. Some key design aspects and performance of single tactile pin is evaluated and reported.

Future work in the project includes:

i) Complete development of foot display to the proposed design in Figure 3.
ii) Development and exploration of various virtual terrains.
iii) Incorporation of existing prototype for programmable foot- pedals [10] to ensure fuller haptic feedback.
iv) Development of the pressure control unit of display to include tactile sensation rendering in virtual environment.

Bibliography

Abulgasem, A. I., (2004) Virtual Simulator for Walking, *Postgraduate Dissertation*, University of Salford.

Caldwell, D. G., (2005) Haptic Displays and Applications, *Positional Notes*, University of Salford.

Caldwell, D. G., Tsagarakis, N. and Gesler, C., (1999) An Integrated Tactile/ Shear feedback Array for Simulation of Finger Mechano receptor, *Proceedings of IEE ICRA,* Michigan.

Fipaza, M. S., (2005) Locomotion Interface in Virtual Environment, *IRIS Conference, Abstract,* Manchester.

Fisch, A., Mavroidis, C., Bar-Cohen, Y. and Melli-Huber., (2005) Haptic Devices for Virtual Reality, Telepresence and Human Assistive-Robotics, *Chapter 4*, 2005.

Kennedy, P. M. and Inglis, J. T., (2002) Distribution and behaviour of glabrous cutaneous receptors in the human foot sole, *Journal of Physiology*.

Masako, D., Mochimaru. M and Kouchi, M., (2003) The Tactile Sensitivity and Elasticity of the Sole of the Foot as a Factor of the Shoe Comfort., *CREST, JST*.

Sarakaglou, I., Tsagarakis, N. and Caldwell, D. G., (2005) A portable Fingertip Tacticle Feedback Array – Transmission System, Reliability and Modelling, *EuroHaptics*.

Schmidt, H., Hesse, S., Bernhardt, R. and Kruger, J., (2005) Haptic Walker….A novel haptic foot device, *ACM Transaction of* A*pplied Perception (TAP)*.

Taylor, P. M., Hosseini-Sianaki, A. and Varley., C. J., (1996) An Eletro –Rheological fluid based tactile array for virtual environments, *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*, Minnesota.

Performance Assessment of a 3 DOF Differential Based Waist Joint for the "iCub" Baby Humanoid Robot

W M. Hinojosa, N G Tsagarakis, G Metta, F Becchi, J Sandini and D G Caldwell

Centre for Advanced Robotics and Automation

Introduction

Anthropomorphic type robots combine many desirable features such as natural human like locomotion and human friendly design and behavior. As a result of this multi degree of freedom human like robots have become more and more common and many humanoid robots have recently been designed and fabricated.

The first biped humanoid robot was WABOT developed at Waseda University in 1973. This biped which was able to perform simple static walking was followed by the development of WABIAN I and II. WABIN-RII. Following these first prototypes a number of other human like robots were developed including the H6, H7 at the University of Tokyo (Nishiwaki et al. 2000), the impressive humanoid robots P2, P3 and ASIMO developed by HONDA (Hirai 1997; Hirai et al 1998; Hirose et al. 2001) and some more recent prototypes such as the JOHNNIE the anthropomorphic autonomous biped robot (Gienger, Loffler and Pfeiffer 2001) developed at University of Munich and the HRP, HRP-2 developed by METI in Japan (Inoue et al. 2001; Kaneko et al. 2002; Kaneko et al. 2004; Kanehira et al. 2002).

Other less well know medium and small size humanoids include SAIKA (Shirata, Konno and Uchiyama 2004) and KENTA Mizuuchi et al. 2002), the MK.5 a compact size humanoid robot constructed by Aoyma Gakuin University (Furuta, Okomura and Tomiyama 2000), the PINO platform constructed by ERATO (Yamasaki et al. 2000) and the SDR-3X (Sony Dream Robot-3X) and SDR-4X developed mainly for entertainment (Fujita 2003; Kuroki et al. 2001).

In the above examples the waist joint is usually implemented using a simple serial mechanism with 2 DOF. This waist configuration is adequate for these robots since the tasks that are usually performed are limited to entertainment and amusement applications and in demonstrating walking capabilities.

The concept behind the development of iCub is to provide the cognition research community with an open architecture human like hardware/software platform for understanding of cognitive systems through the study of cognitive development. The iCUB will replicate a 2 and a half year old human baby acting in cognitive scenarios, performing tasks useful to learning and interacting with the environment and humans. In the early definition of the project two main tasks were considered from which the design requirements for the waist mechanism were derived. These are crawling and manipulation (Metta, Vernon and Sandini). Based on the requirements implied by these two tasks the design of the waist mechanism of the iCub was realised. A 3 DOF differential based mechanism was employed to provide not only increased stiffness

but also increased range and flexibility of motion for the upper body.

This paper presents the design and control of this differential based waist mechanism of the iCub. The paper is organised as follows: Section II gives the general specifications of the waist joint in terms of DOF, range of motions and torque requirements. The following section describes the mechanical design adopted for the waist mechanism and highlights the advantages of this approach. Section IV introduces the system model used for the design of the control scheme while section V presents the control system design. Estimated performance measures of the control scheme are presented in sections VI and VII, by means of simulation and experimental result. Finally, section VIII introduces the conclusions of this work.

Waist Specifications

The kinematics specifications of the waist joint of the iCub include the definition of the number of D.O.F required and their actual location as well as the range of motions. These were defined with attention given to address the requirement for crawling and manipulation and in general to imitate the human baby form. As has been mentioned the iCub will have the approximate size of a two and a half year old child (Metta, Vernon and Sandini). The D.O.F required for the waist was determined by considering both crawling and manipulation scenarios. Crawling simulation analysis showed that for effective crawling a 3 D.O.F waist is essential, Table I.

	Degrees of Freedom (°)	
Joint	Human	iCub
Waist	3 Roll Pitch Yaw	3 Roll Pitch Yaw
		=3DOF

Table 1: Waist mechanism number of DOFS

An additional advantage that a 3 D.O.F waist will offer is the increased range and flexibility of motion for the upper body. This increased flexibility results in an amplified workspace of the iCub when performing manipulation tasks using its hands when in a sitting position. As manipulation is directly related to learning which is an essential task for the iCub the 3 D.O.F waist will provide significant benefits. Based on the above the iCub waist needs to provide 3 D.O.F enabling pitch, roll and yaw of the upper body.

As the iCub is a human-like robot and will perform tasks similar to those performed by a human, the range of motions of a standard human were used as a starting point for the selection of the movable ranges for the waist joints of the iCub.

Table 2 below introduces the range of motion specifications for the joints of the waist mechanism in comparison with the corresponding ranges found in the human.

Waist	Range of motion (°)	
Joint	Human	iCub
Waist roll	-35, +35	-90,+90
Waist pitch	-30, +70	-10,+90
Waist yaw	-40, +40	-60,+60

Table 2: rangle of motion of the waist joint

Looking at Table 2 it can be observed that the range of motion in some joints has been extended or modified. In particular, the range of the waist yaw and roll has been increased while the range of the pitch motion was modified to increase the upper body forward angle to improve the front workspace of the robot. This extends the area where the iCub can reach and manipulate objects.

The last specifications to be discussed are the required torques for the waist mechanism. This forms the starting point for the selection of the actuation groups. To optimise the selection of actuators and reduction ratios, iterations of the mechanical design of the waist and simulated analysis of the system were carried out.

The selection of the type of actuator to power the waist of the iCub involved simulations of the robot while performing crawling motions with different speeds and transitions from a sitting to crawling pose and vice versa. From these simulations, the peak torque requirements of each joint of the waist mechanism were identified as presented in Table III.

Waist joint	Torque Required(Nm)
Roll	30.1
Pitch	45.8
Yaw	27.2

Table 3: Torque required for the waist mechanism

Waist Mechanical Design

The CAD model and the first prototype of the waist mechanism of the iCub baby humanoid robot are shown in Figure 1 and 2. The iCub waist was designed using the 3D CAD software *Pro Engineer Wildfire 2 from PTC.*

In the previous section, the role of the waist joint in the flexibility of motion of the upper body was highlighted. Such flexibility must be accompanied by high positional accuracy of the upper body as is required during manipulation.

For the pitch motion of the waist, the two actuator groups that power the pitch and yaw motion apply a synchronous motion to the two differential input wheels using the torque of both motors. For the yaw motion, the motors turn in opposing directions. This causes a yaw motion on the upper body, again using the torque of both motors.

This differential mechanism has several advantages when compared with traditional serial mechanisms used in humanoid robots. These are:

i) Increased stiffness and accuracy.

ii) The sum of the torque generated by the two actuators that power the differential joint can be distributed in both joints.
iii) As a result, smaller actuators can be used to achieve the maximum output torques required for the pitch and yaw motions.

The roll motion is implemented with the incorporation of a pulley shaft that is directly connected to the upper body frame. Torque is conveyed through a cable transmission system that provides also additional gearing to meet the torque requirements of the roll joint, Table III. For the first prototype DC motor actuators were employed to power the waist joints.

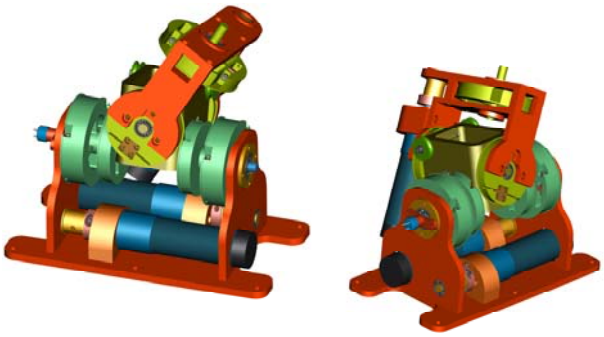


Figure 1: CAD captures of the differential drive to be used as the robot waist.

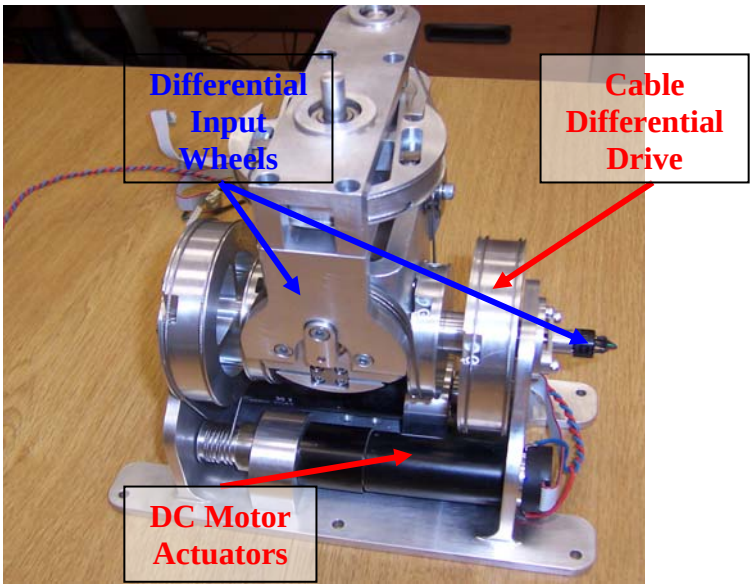


Figure 2. Prototype of the differential based waist mechanism.

System Model

In order to achieve effective control of waist motions an accurate model of the system has been developed and extensive simulations performed.

Motor Actuator Model

For the development of this system model, a model of the motor actuator used to power the waist joints was required. Equations (1) and (2) characterise a general DC motor.

$$J\frac{d^2\theta}{dt^2} = T - B\frac{d\theta}{dt} \Rightarrow \frac{d^2\theta}{dt^2} = \frac{1}{J}\left(K_t i - B\frac{d\theta}{dt}\right)$$

(1)

$$L\frac{di}{dt} = V - R \times i - e \Rightarrow \frac{di}{dt} = \frac{1}{L}\left(V - R \times i - K_e \frac{d\theta}{dt}\right)$$

(2)

This mathematical model was implemented using MATLAB Simulink, Figure 3, and the results were referenced with the real system.

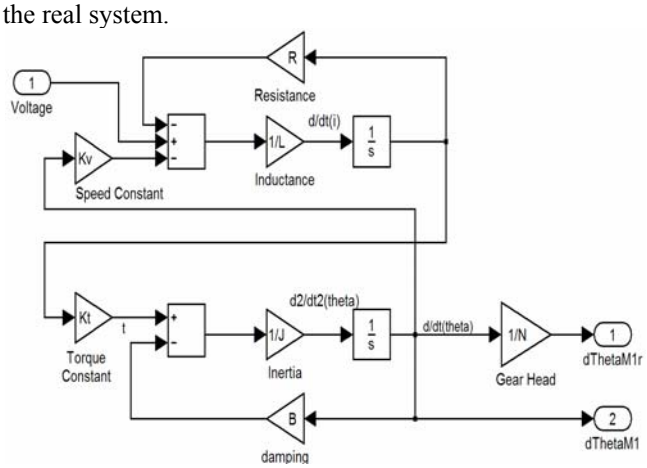


Figure 3: MATLAB Simulink representation of the mathematical model of the two DC motors used to drive the differential waist mechanism.

Differential Mechanism Model

There are 2 motors controlling the pitch and roll motion of the robot waist, both located in parallel one in front of the other, but with its shafts pointing in opposite directions. From the kinematics analysis of the joints, the differential equations for the differential driver are:

$$R \times \left(\theta_{M1} + \theta_{M2}\right) = \theta_{roll} \tag{3}$$

$$R \times \left(\theta_{M1} - \theta_{M2}\right) = \theta_{pitch} \tag{4}$$

Where R is the gear head reduction ratio for motor 1 and 2, θ_{M1} and θ_{M2} are the rotor angles for motor 1 and 2 respectively and θ_{roll} and θ_{pitch} are the waist roll and pitch rotation angles.

Waist Dynamics

In order to have an accurate mathematical description of the system the dynamics of the system must be included in the simulation; this includes the weights of the mechanical assembly, friction of the contact joints and motors and the inertia of the system. In Figure 4 the block named “spine dynamics” includes the waist kinematics (combines the motion of the two DC motors in differential mode) and the dynamics of a generic limb with a variable weight used to perform a wide range of tests under different sets of conditions.

Complete System Modelling

In the simulations performed, the position feedback is calculated by integrating the speed output from the DC motor model, in the real system the speed is calculated using the derivative of the position output acquired from incremental encoder

readings. Figure 4 shows the final control scheme including the differential equations block and the system dynamics for both motors, for a 2 DOF motion.

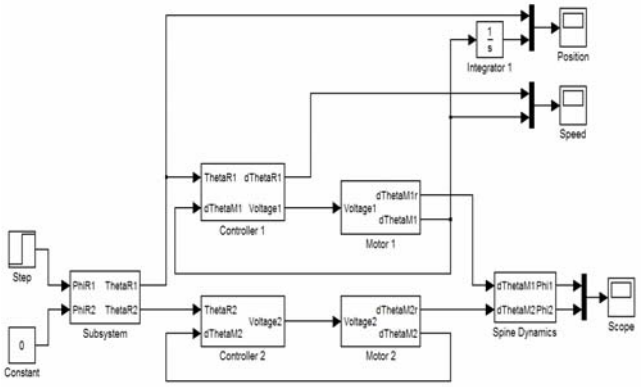


Figure 4: Complete system for the 2 DOF differential joint. Speed and position control of 2 DC motors configured as a differential drive.

Control System

Many different approaches exist for control, each one with advantages and disadvantages, but the classic PID control scheme is still implemented in about 90% of real systems. In this section, the design, model and simulation of a dual loop PID control system with a dynamic anti-windup scheme will be described.

A traditional PID controller uses the derivative and integral of the error signal; but when the reference input changes, the tracking error changes rapidly. The derivative signal of the tracking error can generate shock on the system (Astrom 2005). To avoid this, the controller proposed uses the derivative of the output signal as shown in Figure 5.

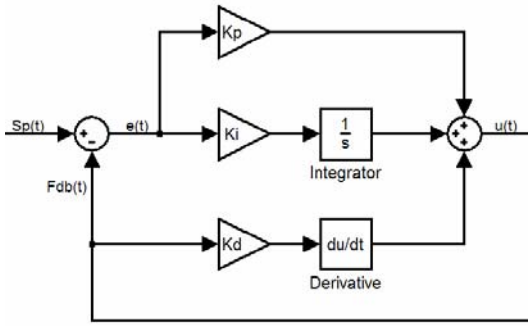


Figure 5: Block Diagram of the classic three-term control scheme with derivative term using the feedback signal.

In order to perform position and speed control, a cascade controller was implemented. As observed in Figure 6 a first inner loop is used to control speed, this loop uses the position controller (outer loop) output as a reference signal. The speed is then established by setting the maximum and minimum values of the output limiter of the position controller. The fact that the speed is to be adjusted by setting a maximum and a minimum value to the limiter implies that the system will saturate. A simpler limiter will limit the position control output to a maximum value, which represents the desired maximum rotor speed. If this speed is low enough, the rotor can take a considerable amount of time to reach its final position, which in turn will produce a high integral term value.

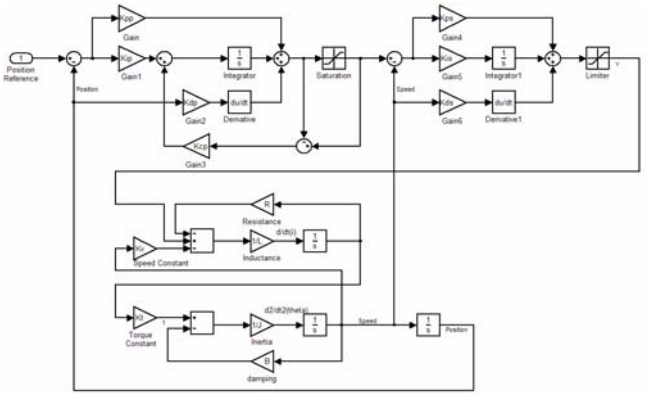


Figure 6: Cascade control scheme for position and speed control, including the DC motor model.

In order to avoid this, a dynamic anti-windup scheme (Astrom 2005) was implemented; this scheme, Figure 7, takes the output of the system and compares it with a pre-saturated output to obtain a saturated error value, this value is then multiplied by a previously set gain and then introduced to the integrator, reducing its value. Using this scheme it is possible to ensure that the integral term will not have a high value so when the signal reaches its set point it will effectively reduce the steady state error, further more reducing the overshoot. Figure 6 shows the mentioned scheme.

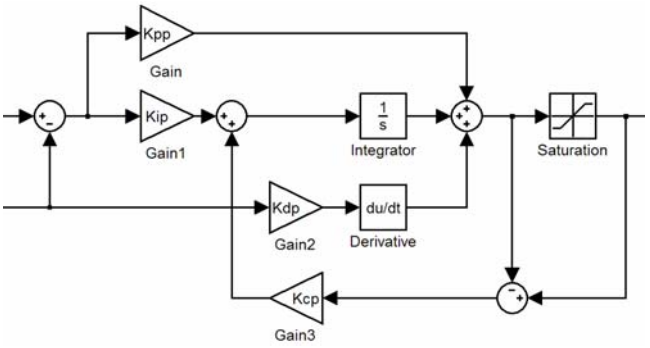


Figure 7: Controller with dynamic anti-windup scheme, the integral term is effectively reduced, preventing overshoot and the saturation of the integral term.

Simulation Results

Testing of the control schemes were carefully evaluated and rigorously analyzed through simulations. For the following simulations, 2 types of reference signal where used: a step input signal and a sinusoidal signal at a range of different frequencies. In all cases, the solid lines represent the reference signals and the dashed lines the rotor position. In the following graphs, the only load considered was the weight of the limbs; the frictions in the joints were not considered.

Speed Control

The balance of the body of the robot is greatly affected by the speed of motion due to inertial forces generated by the body weight. The following graphs show the simulated results of the speed controller, with a forced reference input signal (the actual speed reference is generated by the position controller). These simulations were performed using MATLAB Simulink. The actuator groups used for this test consist of two MAXON DC motors and gearboxes with a maximum output speed after the reduction of 440 deg/s.

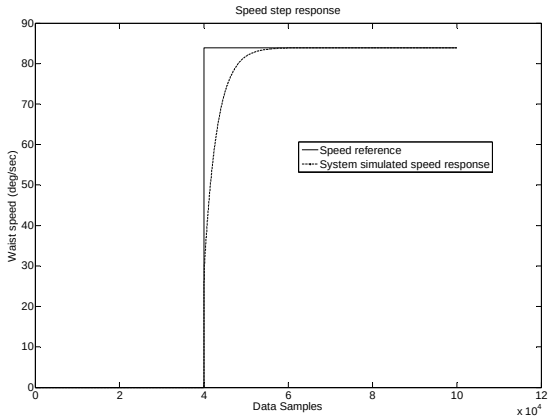


Figure 8: Step response of the motor speed controller without load. This will help us evaluate the motor acceleration capabilities and the controller error margins.

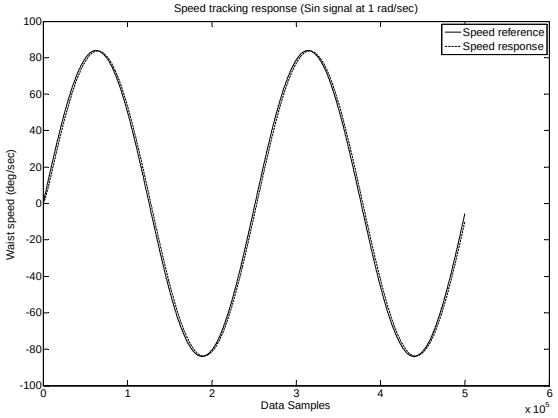


Figure 9: With a 1 rad/s sinusoidal reference and a 1kg load over the joint.

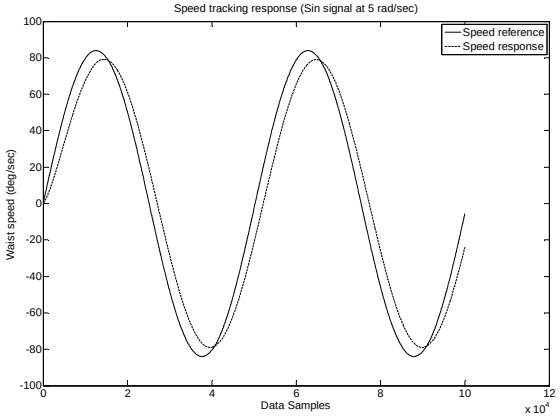


Figure 10: With a 5 rad/s sinusoidal reference.

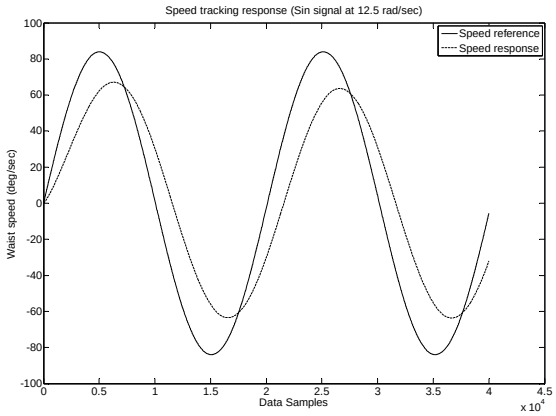


Figure 11: With a 10 rad/s sinusoidal reference.

For the simulations results presented, reference signals of 84 deg/s (1.46 rad/s) were used and these represent the rotor speed for one of the motors.

This allowed the evaluation of the motor speed control capabilities under loaded and unloaded conditions and different motion directions (the load presented to the motor changes as the position of the centre of gravity of the robot's body moves)

The simulation results shows a good speed response time capable of moving the body of the robot at acceptable speeds.

Position Control

The following graphs show the simulation results of the position using the complete model (position and speed controller cascaded).

The effect of a 1kg load in the tracking capabilities of the position controller can be seen; the load slows down the motion of the joint, even with a speed set near the motor maximum speed, though the position controller was able to position the waist joints with errors around 0.03deg in steady state.

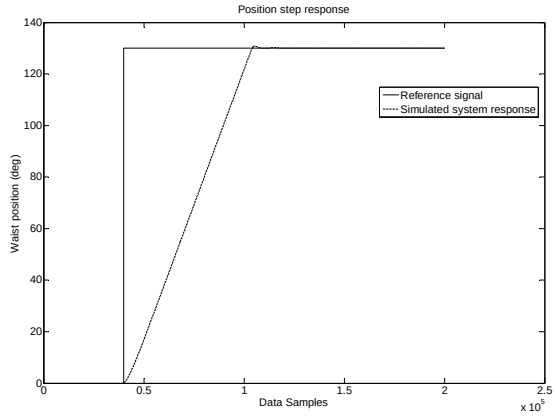


Figure 12: Response for a square signal reference.

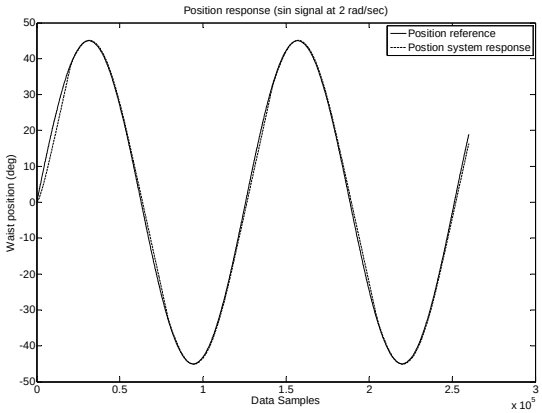


Figure 13: Position tracking response with a 2 rad/s sinusoidal reference with a 90deg. of motion range.

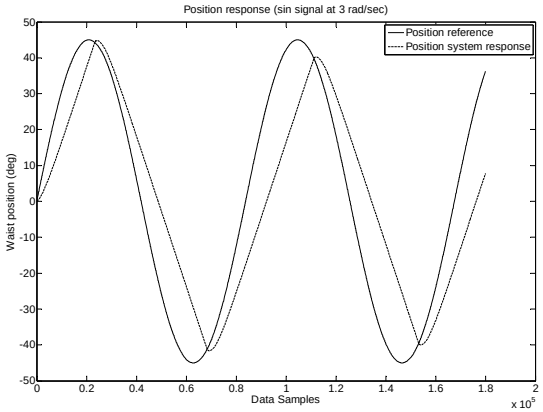


Figure 14: Position tracking response with a 3 rad/s sinusoidal signal and 90deg of motion range.

Due to mechanical limitations, it can be observed from Figure 14 that the robot waist was not able to follow the reference signal accurately at rates of change higher than 2 rad/s even when setting the speed limit at its maximum.

System Experimental Results

The following results were obtained by collecting measurements from the real system through the microcontroller (TMS320F2810, DSP from Texas Instruments) on which the control system was implemented, through its JTAG interface to a PC. Figure 15-17, show the results for the position control, when a step input is presented, with a set speed of 84 deg/s in the joint. The joint is set to move forward and backwards between 22 and 65 deg.

After evaluating the results from the real system, different factors not considered in the simulation model, like friction in the waist joints, were observed to have little effect on the results.

From the graphs below, errors of under 0.05deg in the position and under 1 deg/s for the speed can be observed.

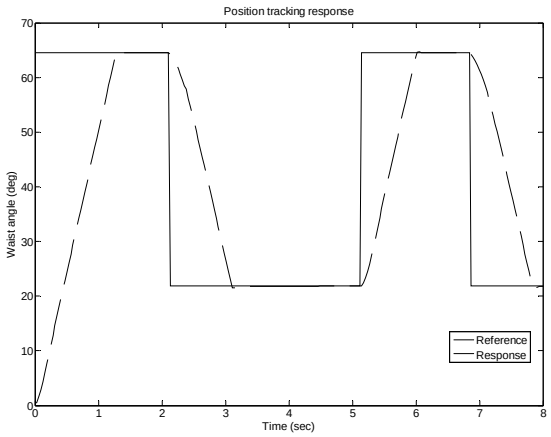


Figure 15: Position Tracking Response of the real system. Reference signal in solid line, actual waist position in dashed line.

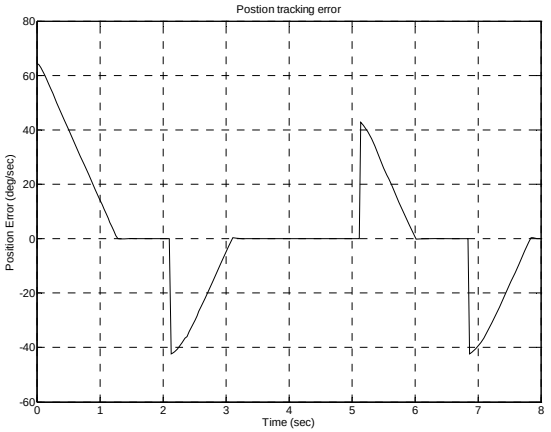


Figure 16: Position tracking error of the real system.

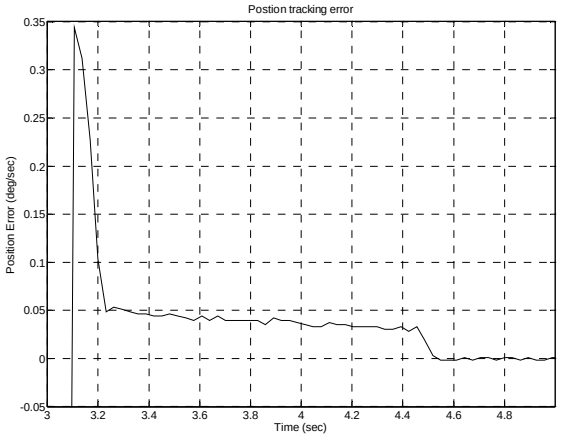


Figure 17: Position tracking error. In this close up it is possible to observe the position error in steady state. Actual results show errors under 0.05deg in steady state with 0.35deg overshoot.

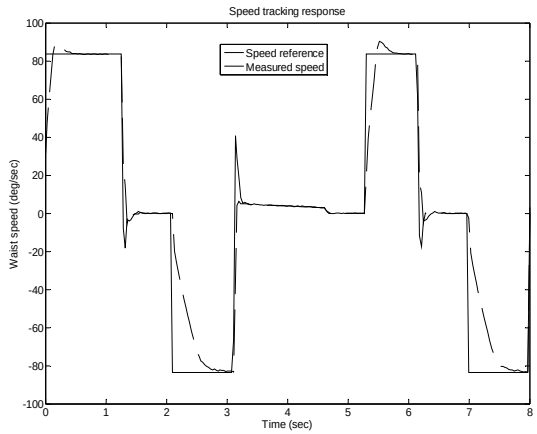


Figure. 18: Speed tracking response of the real system. Reference signal in solid line, actual waist speed in dashed line.

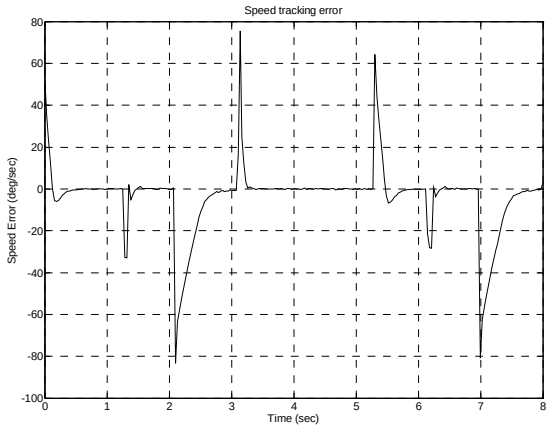


Figure. 19. Speed tracking error of the real system.

Conclusions and Future Work

This work presented the design of a differential based mechanism developed to form the waist joint of a baby humanoid robot. A cascade PID based position and speed controller was developed and its characteristics, such as overshoot, settling time and steady state error, have been evaluated through both experimentation and simulation. A control system consisting on a PID controller was established to achieve accurate position control of the joints.

It has been demonstrated through experimental implementation that the proposed control system can achieve control accuracy of 0.05 deg in step responses. In addition, a favorable speed control for sinusoidal and step trajectories was achieved. The control results presented in this study demonstrate that the proposed mechanism and control system can offer the desired motion range with high positional accuracy.

Future work will include a performance evaluation of the system using variable length and weight bodies to evaluate the effect of inertia on the system as well as mechanical (fatigue, maximum torques, etc.), electronic (current consumption, noise, etc.) and thermal effects.

Acknowledgements
This work is supported by the European Commission FP6, Project IST-004370. Special thanks also to Prof. D. G. Caldwell and Dr. N. Tsagarakis for their continuous support.

References
Åström, K. J., (2005) Advanced PID Control, *ISA – The Instrumentation, Systems, and Automation Society*, 2005.
Fujita, M., Kuroki, Y., Ishida, T. and Doi, T., (2003) Autonomous Behavior Control Architecture of Entertainment Humanoid Robot SDR-4X, *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 960–967.
Furuta, T., Okomura, Y. and Tomiyama, K. (2000) Design and Construction of a Series of Compact Humanoid Robots and Development of Biped Walk Control Strategies, *Proc. IEEE-RAS Int. Conf on Humanoid Robots*.
Gienger, M., Löffler, K. and Pfeiffer, P., (2001) Towards the Design of Biped Jogging Robot, *Proc IEEE ICRA*, pp. 4140-4145.
Hirai, K., (1997) Current and Future Perspective of Honda Humanoid Robot, *Proc. IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, pp. 500-508.
Hirai, K., Hirose, M., Haikawa, Y. and Takenaka, T., (1998) The Development of Honda Humanoid Robot, *Proc. IEEE ICRA*, pp. 1321-1326.
Hirose, M., Haikawa, Y., Takenaka, T. and Hirai, K., (2001) Development of Humanoid Robot ASIMO, *Proc. Int. Conference on Intelligent Robots and Systems, Workshop2* (Oct.29, 2001).
Inoue, H., Tachi, S., Nakamura, Y., Hirai, K., Ohyu, N., Hirai, S., Tanie, K., Yokoi, K. and Hirukawa, H., (2001) *Overview of Humanoid Robotics Project of METI, Proc. the 32nd Int. Symposium on Robotics*, pp. 1478-1482.
Kanehira, N., Kawasaki, T., Ota, S., Akachi, K., Isozumi, T., Kanehiro, F., Kaneko, K. and Kajita, S., (2002) Design and Experiment of Advanced Leg Module (HRP-2L) for Humanoid Robot (HRP-2) Development, *Proc. Int. Conference onIntelligent Robots and Systems*.
Kaneko, K., Kanehiro, F., Kajita, S., Yokoyama, K., Akachi, K., Kawasaki, T., Ota, S. and Isozumi, T., (2002) Design of Prototype Humanoid Robotics Platform for HRP, *Proc. IEEE/RSJ Int. Conference on Intelligent Robots and Systems*, pp. 2431-2436.
Kaneko, K., Kanehiro, F., Kajita, S., Hirukawa, H., Kavasaki, T., Hirata, M., Akachi, K. and Isozumi, T., (2004) Humanoid Robot HRP-2, *Proceedings of the 2004 IEEE ICRA, New Orleans, LA*.
Kuroki, Y., Ishida, T., Yamaguchi, J., Fujita, M. and Doi, T., (2001) A Small Biped Entertainment Robot, *Proc. IEEE-RAS Int.Conference on Humanoid Robots*, pp. 181-186.
Metta, G., Vernon, D. and Sandini, G., D8.1 Initial Specification of the CUB Open System, *http://www.robotcub.org*.
Mizuuchi, I., Tajima, R., Yoshikai, T., Sato, D., Nagashima, K., Inaba, M., Kuniyoshi, Y. and Inoue, D., (2002) The Design and Control of the Flexible Spine of a Fully Tendon-Driven Humanoid "Kenta", *IEEE/RSJ IROS*, pp. 2527-2532.
Nishiwaki, K., Sugihara, T., Kagami, S., Kanehiro, F. and Inaba, M., (2000) Design and Development of Research Platform for Perception-Action Integration in Humanoid Robot: H6, *Proc. Int. Conference on Intelligent Robots and Systems*, pp. 1559-1564.
Shirata, S., Konno, A. and Uchiyama, M., (2004) Design and Development of a Light-Weight Biped Humanoid Robot Saika-4, *Proc. IEEE/RSJ IROS*, pp. 148-153.

Yamaguchi, J., Soga, E., Inoue, S. and Takanishi, SA., (1999) Development of a Bipedal Humanoid Robot – Control Method of Whole Body Cooperative Dynamic Biped Walking, *Proc. IEEE Int. Conference on Robotics and Automation*, pp. 368-374.
Yamasaki, F., Matsui, T., Miyashita, T. and Kitano, H., (2000) PINO The Humanoid that Walks, *Int.Conference on Humanoid Robots*, 80.pdf.

The Source of Artistic Impression: An Overview of the Open Source Practice in the Art and Design Context

Vivien Wei Liang

Adelphi Research Institute for Creative Arts and Sciences

Introduction

Open Source Software is the most prominent example of open source. Its principle concerns the availability of software source code as open source under a specific license to study, change and improve its design (www.wikipedia.org). Beyond the software development context, Open Source has been regarded by some as a philosophy or a pragmatic methodology (Mulgan, Steinberg, Salem 2005 & Naughton 1999).

While the concept and ideology of Open Source has become widely spread and many projects have been generated based on the term of "Open Source", it is nevertheless that the actual method that has not yet been formally addressed in the conceptual aspect. According to McConnell (1999) "much of the Open Source Methodology resides primarily in the informal legend, myth and lore surrounding specific projects like Linux".

In the creative art context, the practices of the arts have shown some similarities to Open Source. These can be seen in the early works before the term was known by people in the field. Examples of this can be seen particularly in the areas of Public Arts and New Media Arts, which where many artistic works or creative projects were created based on the element of openness and collaborative participation.

Although the definition of Open Source in the arts field has not yet been formalised, Lovejoy (2004) has the closest identification of the concept so far. She describes 'Open System' Art works as fostering one's power to participate in a web of connected communities, and a place where negotiation takes place as a form of shared authorship and social exchange.

Open Source Application in Creative Arts

"Technological advancement has not only impacted on the way people see and think, it also affects the content, philosophy and style of art works" (Lovejoy 2004).

Artists first began to be concerned and be publicly involved with the use of digital technology in the post war period. Examples can be seen in some artist's works such as John Cage, Ray Johnson, Roy Ascott, who created works and developed ideas that reflected the issue of information and communications technology (Gere 2002).

The rise of the Internet has enabled artists to look in further into the subject of creative networking, which involves exploration of community-building, and cross-pollination of user interaction in the virtual space (Couey 1991). But it is the establishment of Linux that has taken off the whole 'open' issue. The success of the

software project is the icon and the ideology for artists who feel strongly about the principle (Ippolito 2004, Schultz 2005, Sollfrank 2005), and subsequently, the Open Source Phenomenon has lead to a culture and political movement, which artists have followed as a trend. This is expressed in the form of campaigning for the artistic freedom and debating about the legal issues that are either seen as a barrier or protection for artistic works (Kimbell 2004).

On the other hand, the case of Linux has also been seen as an example of an alternative practice for generating creative arts projects. Mulgan, Steinberg and Salem (2005) point out that the Open Source method is a new way of creating knowledge and producing work. As such, there are many organisations or academic institutes like RIXC, Eyebeam, Interspace Media Art Institute that are attempting to adopt the Open Source method as a new form for research and development practice in the creative arts field.

Examples of Open Source in Creative Arts

The politics of technology and the issue of ownership and control of media production and distribution have brought together artists who work with the digital technology. They have then worked to form communities that allow them to gather together and discuss issues (Kimbell 2004). Festivals such as Read Me, V2, PixelACHE, or BEK are the main examples that host annual events for artists and developers working with free/libre and open source software. Some of the events like Piksel (www.piksel.no) aim to focus on the open source movement as a strategy for regaining artistic control of the technology, but also a means to bring attention to the close connections between art, politics, technology and economy (www.pixelache.ac).

Artists' freedom still remains as a challenging issue and much debatable topic within the arts communities. But with the legal provision of the Creative Commons licence, which is devoted to expanding the range of creative work available for others to legally to build upon and share (http://en.wikipedia.org/wiki), collaborative composition of all kinds has become more widespread (Mulgan, Steinberg, and Salem 2005).Online art projects like Telematics Timeline (http://telematic.walkerart. org/timeline/index.html) and OSCULTURE (www.osculture.de) are examples of art projects focusing on Open Source.

In other creative arts areas, academic and research development organisations have been focusing on Open Source as the new method for research practice and for producing innovative works. Workshops and seminars have been set up to fulfil this purpose. RAM (Re-Approaching New Media) is a series of workshops hosted between several institutes across Europe and aims to explore the issue of diverse media culture initiatives, which include the use of open source in contemporary practice of architecture, art and media (http://www.rixc.lv/ram5/). Other examples like the 'Grow your own media lab' (GYOML) project, not only discuss and debate Open Source issues but also manage to make practical achievements. Its purpose is also to assist in the development of creative and open access IT centres across the United Kingdom. The project works with arts centres, artists groups and local people, and aims to generate digital creativity by using recycled computer systems (http://www.folly.co.uk/gyoml.htm).

Discussion

This research has sought to explore the multidisciplinary aspect of Open Source theory and practice. It is intended to investigate what the perceptions are of Open Source in the non-software development context, and to question whether the method could be used in other areas.

The limitation of the current research on Open Source at the conceptual level means that an overview of the current Open Source practice in the case of the creative arts field can assist to providing a clearer understanding on the subject and facilitate further development of the theory.

In the current digital society, cross disciplinary research is one of the main focuses for academia, as the artists will continuously seek inspiration and to generate creativity through technologies. The rise of Open Source not only allows artists to reflect the way they practice creative art works, but it could also be an alternative method for developing creative projects. However, the definition of the Open Source method still remains a question that needs to be answered.

References

Bonaccorsi, A. & Rossi, C., (2003) Why Open Source Software can succeed, *Elsevier Science Journal, Research Policy* 32, p. 1243-1258.

Couey, A., (1991) Art Works as Organic Communications Systems, *Leonardo*.

Folly- Grow Your Own Media Lab Project, (2005) http://www.folly.co.uk/gyoml.htm.

Gere., C., (2002) *Digital Culture*. London: Reaction Books, pp. 75-76

Ippolito, J., (2004) *Why Art Should Be Free*. http://three.org/ippolito/home.html.

Kimbell., L., (2004) *New Media Art: Practice and Context in the UK 1994-20047*. Manchester: Cornerhouse Publications.

Lovejoy., M., (2004) *Digital Currents: Arts in the Electronic* Age. London: Routledge.

McConnell., S., (1999) *Open Source Methodology: Ready for Prime Time*. IEEE Software http://turnbull.sk.tsukuba.ac.jp/Teach/InfoEcon/McConnell-Prime Time.pdf.

Mulgan, G., Steinberg, T. and Salem, O., (2005) *Wide Open- Open Source Methods and Their Future Potential*. London: Demos Publication.

Naughton., J., (1999) *A Brief History of the Future- The origins of the Internet*. London: Weidenfeld and Nicholson.

OSCULTURE (1999) www.osculture.de.

Piksel-Free/Libre and Open Source Audiovisual Software and Art (1999) http://www.piksel.no/.

PixelACHE Festival of Electronic Art and Subcultures (2005) http://www.pixelache.ac/2005/.

RAM05 (2005) *Open Source Media Architecture Workshop*. http://www.rixc.lv/ram5.

Schultz, P., (2005) Current Models and Future Possibilities for Independent Media Production, From Tomorrow on: *Art, Activism, Technology and the Future of Independent Media Conference*. Manchester Conerhouse, 27th October 2005.

Sollfrank, C., (2005) http://www.artwarez.org/.

Telematics Timeline Project (2006) http://telematic.walkerart.org/timeline/index.html

Evidence-Based Practice in Social Work and Child Protection: a Critique of Experimentalism

Guido van de Luitgaarden
Institute for Health and Social Care Research

Introduction
In social work, risks are abundant both for service users, professionals and agencies. Bad things can happen to clients, as can be illustrated by many deaths of children at the hands of their parents, who were under some kind of social work supervision. When bad things happen to clients, responsibility is often placed with social workers and their agencies (Jones and Gupta 1998; Kelly and Milner 1996; Spratt 2000; Webb 2002). In turn, they will be held accountable for what they did or did not do to prevent such things from happening. Social work has to deal with multiple accountabilities. Service users, colleagues, government agencies, and the general public all demand answers whenever something goes dramatically wrong. Auditing (Munro 2004), legalistic preoccupations, and managerial control (Rzepnicki and Johnson 2005) are amongst the means to reduce the risk of being blamed of virtually anything, ranging from a client's violent death to inefficient use of resources.

To minimise the risk of being accused of bad practice, social work is trying to decrease the likelihood that something will go wrong. One of the ways in which this is attempted, is by means of looking for increased certainty through the use of findings from social science in professional practice. The application of 'hard' research evidence to social work practice should minimise these hazards. Evidence-Based Practice is a concept that strives to do such. Having said this it should be clear that, although protecting social workers against outside blame is one aspect of EBP, there are also more intrinsic ethical and moral arguments regarding the well-being of clients which, amongst others, will be dealt with in the course of this paper. This paper aims to show that EBP is not a singular term, and that there are serious problems regarding the application of a strictly experimental approach to EBP in social work practice.

First, the nature and scope of evidence-based practice, as well as the two competing interpretations of what EBP is will be explained. Next, a theoretically oriented critique of experimentalism in social work research will be provided. In the subsequent part of this paper, it will be argued that the ethical, political and moral arguments for basing practice on evidence are connected to the theoretical underpinnings of experimentalist EBP. As these underpinnings are usually not valid for social work research, the ethical aspects are directly influenced. As a result, some of these ethical arguments will be shown impossible to hold. It is concluded that social work is better off with a pragmatic interpretation of what EBP should entail,than with a strictly experimentalist approach.

Nature and Scope of EBP
To get a grasp of the nature and scope of EBP, its definitions and process will need to be dealt with first. The two main interpretations of what EBP *is* or *should be* are addressed in conjunction with a critical analysis of their epistemological underpinnings.

Definitions

Sackett et al. (1996) first proposed the concept of Evidence-Based Practice in the medical field. They defined evidence-based medicine as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients". A year later, the same authors broaden the definition through the change of the word 'patients' into 'individuals' and changing 'Evidence-based medicine' into 'Evidence-based practice' (Sackett et al. 1997), thus suggesting that the concept might not be exclusively used in the medical field.

Other definitions have been proposed, such as "Intervention based on the best available evidence" (McNeece and Thyer 2004), "In its most basic definition, [EBP] aims to provide the best available evidence at the point of [contact with the client] (Gibbs and Gambrill 2002). These definitions seem to paraphrase Sackett et al.'s definition. Among social work scholars widespread consensus exists about the usefulness of Sackett et al.'s definition (Ainsworth and Hansen 2002; Gambrill 1999; Garretsen, Rodenburg, and Bongers 2003; Gilgun 2005; Gira, Kessler, and Poertner 2004; MacDonald 1999; MacDonald and Sheldon 1998; Melief 2003; Murphy and McDonald 2004; Newman and McNeish 2002; Sheldon and Chilvers 2000; Taylor and White 2001). However, because the definition is rather general and broad, it needs to be interpreted for use in practice. Whenever an attempt is made to define the constituent parts of the definition, heated academic debate and even polemics are often the result. Views differ profoundly when it comes to questions as to what can be considered 'evidence', '*best* evidence', and what 'the conscientious, explicit, and judicious use' thereof is. Although the definition itself is hardly contested, four areas of controversial issues remain: Epistemological issues, ethical issues, factors that inhibit or facilitate EBP within the practice context, and the relationship between external evidence and experience, expertise and intuition. As stated before, these areas will be examined in the course of this paper.

Process

Like the definition, the views on what the process of evidence-based decision making should look like, is not very contested either. Sackett et al. (1997) describe the process of EBP as follows:

The practitioner, when presented with a case, has to formulate an answerable clinical question. Next, they need to do a search for relevant empirical findings and meta-analyses, and critically appraise the research that is found. This critical evaluation should be done in terms of impact, validity and relevance to the case at hand.

The evidence that has been found will have to be integrated with the expertise of the worker and the values and circumstances of the client (Greenhalgh 1999).This allows the client and the worker to make an informed decision about the desired outcomes and the interventions that are most likely to render those results.

The final step is evaluating the experiences the practitioner has had while conducting the previous steps, and especially the outcomes of the services that have been provided to the client. The term 'outcomes' refers to problem resolution or enhanced functioning of the client.

Those authors that mention the steps involved in EBP, refer to this process description or present a derivate thereof (Gambrill 1999, 2001, 2003; Gibbs and Gambrill 2002; Gilgun 2005; Gira et al. 2004; McNeece and Thyer 2004).

So, practitioners should ask themselves whether, when presented with a case, one intervention, as compared to another intervention or no intervention at all, leads to the desired results.

Interpretations of EBP and their Epistemological Underpinnings

The definition by Sackett et al. (1997), has been interpreted along two lines: A narrow interpretation and a more inclusive one, both of which are established on the basis of differing paradigms. Trinder (2000) calls the former the 'experimental' approach, and the latter the 'pragmatic' approach. These approaches will be discussed here. Attention will be given to the epistemological arguments that are employed to support either approach.

Any interpretation of the definition of EBP departs from a basic system of beliefs on which certain ontological, epistemological, and methodological assumptions are based. In many cases, these assumptions about the nature of social work reality, the relationship between the researcher and reality, and how researchers can come to know that reality, are the source of academic disagreement. Here, these issues will be addressed and critically examined, and it will be argued that a pragmatic, inclusive view of EBP is likely to be more fruitful in the field of child protection than a strictly experimental approach.

Experimental EBP

The experimental approach focuses on 'what works', and prescribes a number of steps that are to be taken in order to gain valid knowledge about the effectiveness of interventions. The first step in gathering evidence about what works, is the formulation of a clinical question. A clinical question consists of the four elements of Problem, Intervention, Comparison and Outcomes. These questions usually take the following form: "In the case of problem A, does intervention B when compared to intervention C lead to an improvement in outcome D?"

These are the kinds of questions that can be answered in different ways. In other words, multiple kinds of evidence resulting from various kinds of research could be used to answer them.

Taxonomy of Evidence

The main distinction between the experimental and pragmatic school of EBP lies in the fact that the former operates a strict hierarchy of evidence in which certain kinds of research are preferred over others.

This theme of what kind of research yields 'best evidence' is one that is subject to theoretical dispute. The taxonomy of scientific evidence in medicine which was proposed by Guyatt (1995) and embraced by some social work scholars (Geddes 2000; McNeece and Thyer, 2004) causes a lot of turmoil as quantitative approaches are regarded as being superior to qualitative enquiry. This taxonomy puts systematic reviews and meta-analyses on top, followed by Randomised Controlled Trials

(RCTs), Quasi-Experimental Studies, Case-Control and Cohort Studies, Pre-Experimental Group Studies, Surveys, and ends with qualitative studies. RCT's are considered to be the 'gold standard' in this approach and reviews of those RCT's are supposed to yield knowledge about 'what works'.

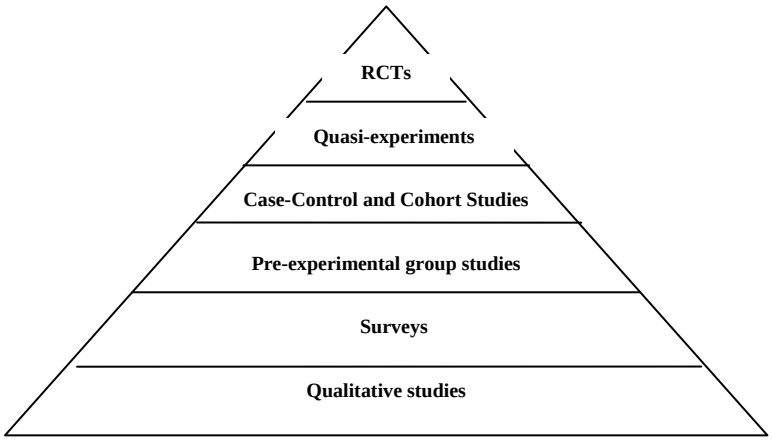


Figure 1: A taxonomy of scientific evidence (Guyatt 1995)

Theoretical Difficulties of Experimentalist EBP in Social Work

Objections to the Taxonomy of Evidence

The experimental approach to EBP is criticised by many, arguing that a narrow view has developed of what can be considered legitimate ways to arrive at practice-relevant knowledge (Butler 2003; Taylor and White 2001).

It has been labelled "positivistic", "utilitarian", "scientistic" (Butler 2003) and "behaviourist" (Humphries 2003). This is where paradigmatic differences of view are introduced in the discourse about EBP. It would make sense to consider paradigmatic issues before leaping to the matter of method because sensible discussion about methodology can only take place whenever there is some kind of agreement on the ontological and epistemological underpinnings of the methodologies in question (Humphries 2003). It is meaningless to argue about ways in which reality can be known if there is no reference to the nature of that reality and the position of the researcher within that reality.

Ontological Arguments

The approach of logical positivism that has been associated with the taxonomy of evidence discussed above, strives to conduct social science along the same lines as the natural and physical sciences (Guba and Lincoln 1994). That is, cognitive-behavioural approaches are preferred, in which behaviour can be observed, measured and then explained rationally in terms of causes and effects. From these operations, general laws are to be established.

As in the natural sciences, the idea goes that if a group of people who are dealing with a certain kind of issue are randomly assigned to either an intervention-group or a control-group that does not receive any kind of intervention, any statistically significant differences in behaviour between those groups can be ascribed to the intervention in question.

Opponents of the experimental school of EBP suggest that when quantitative methods are considered to be as useful as, or more useful then qualitative inquiry in social work, that such constitutes positivism. However, when reviewing the literature on EBP in Social Work, it must be concluded that only Bruce Thyer (1993, 2002) explicitly refers to positivism as a preferred paradigmatic stance.

Although he refers to Logical Positivism, he also remarks that there are limitations to this approach and that certain kinds of questions are not amenable to it, thereby conceding that research by other means is necessary as well, especially for the purpose of generating hypotheses. Thus, it can be concluded that those who argue that logical positivism is the best way to arrive at legitimate social work knowledge hold (at best) a minority position in the world of social work research.

It is argued that positivism in the social sciences has been abandoned dozens of years ago because of the induction problem, which makes a verificationist approach invalid. (Gambrill 1999).

Antagonists of strictly experimental EBP have a tendency to project an image of positivism onto those who embrace the use of RCTs and other quantitative methods and then set out to explain why positivism is a useless approach to social work research (Oakley 2000).

The assumption of positivism that social reality or aspects thereof are 'out there' is countered with the argument that social reality consists of people giving meaning to their own and others' behaviour, and that social reality is a result of language and discourse (Mead 1934; Witkin 1999). Especially in the field of child abuse, it can be argued that child abuse *as such* does not exist in and of itself. Although the act of beating a child is likely to be very real, the *labelling* of that beating as *child abuse,* as well as societal reactions to it, are value judgments that are constructed, codified, reconstructed and changed over space and time. Therefore child abuse as an objective, singular entity simply does not exist as is probably best shown by the fact that notions of what is appropriate and what is inappropriate behaviour towards children differs between (sub)cultures and continually changes over time (Buckley 2003; Corby 1993).

Epistemological Arguments

Similarly, experimental EBP is countered with the epistemological argument that because of the subjective nature of social reality, researchers cannot be detached and value-free. Researchers themselves are human beings as well, and as a consequence they bestow meaning on their own actions and those of others. The way researchers theorise reality, completely determines the shape and contents of the research tools they use for data collection, as well as for subsequent interpretation of those data. In addition, although various forms of physical assault are very real, they cannot generally be observed. They can only be inferred from other information such as doctors' reports, testimonies, etc. which are usually incomplete, ambiguous and contradictory. This makes a behaviourist approach practically impossible in child protection. Thus, experimentation is no less subjective than other methods and cannot be hierarchically above those methods.

Methodological Issues

Specific methodological problems of experimentation in the social and behavioural sciences, are those of standardisation and operationalisation. Standardised measures are not any more valid than other research tools because, as we have argued before, there is nothing to be measured as a result of the unique relationship all of us have with our own realities. Constructed scales and their subsequent validation lead to denaturalisation, encapsulation and objectification of meaning (Burnard and Hannigan 2000; Massé 2000). Thus, approaches that draw on standardisation inherently ontologise social reality. As has been shown before, such is impossible in the field of child protection, as well as most aspects of the social sciences in general. For example, child neglect can have an entirely different meaning to one person than it has to the other. Whereas one might have experienced child neglect as a completely negative situation, another might think of it with mixed feelings, as a lack of parental interest in the child also renders it a sense of freedom. An abstract concept, such as 'neglect' cannot be observed. Because of this, attempts to come to know to what extent a child is neglected are made by operationalising the concept by formulating parameters of abuse and neglect and the gravity of these phenomena (Baeten and Willems 2004). The researcher constructs these parameters, and subjects are then asked to report where they would place themselves or others at such a parameter. Alternatively, a practitioner might 'score' the situation of a family on a decision-aid, which is inherently a value-judgment on the basis of an inference made on the basis of incomplete, contradictory, ambiguous information. Thus, the designer of the research instrument frames the answer beforehand by superimposing a certain definition and parameters of neglect. Furthermore, the problem with self-reported behaviour is that it is prone to a multitude of biases. Because quantification can only take place through inference instead of direct observation, we might be able to roughly approach a socially constructed measure of neglect by quantitative means, but we will never be able to measure it as exactly as we would determine the weight of a bag of sugar.

Methods are nothing more than a means by which we can improve our knowledge and all methods have to deal with issues of validity and reliability, which is something quite different from objectivity (Humphries 2003; Oakley 2000). Contrary to most parts of medicine, in the social sciences quantification and measurement are not the same. Quantification in the social sciences consists of value judgments expressed in a number. Because they are expressed in a number, it does not make them more objective.

It should not be concluded on the basis of the reasons provided above that quantification or experimentation are useless or even inferior to qualitative inquiry. The conclusion to be drawn here is that, as no claims to objectivity can be made, quantitative approaches to social research should be regarded as equal to other methods. This stance is taken by the proponents of a more pragmatic approach to evidence-based practice, which will be explored next.

The Pragmatic Alternative

As has been argued before, it can be concluded that a taxonomy of evidence is hardly possible to maintain in social work research. However, although experimental approaches cannot be regarded as superior to other approaches, this does not mean

that they are necessarily useless. Scholars who are critical of experimental EBP propose an alternative, pragmatic stance towards knowledge for practice.

The need to do so arose from the experienced unproductiveness of discussions about quantitative versus qualitative research that have obstructed social work innovation (Gibbons 2001; Humphries 2003; Newman and McNeish 2002; Tierney 2002).

Critical Realists argue that, although it is believed that there is a reality 'out there' (Taylor and White 2001) we cannot come to know that reality in an absolute sense. What social reality would then look like, regardless of the interpretations of people, is a question that remains unanswered. This ontological shortcoming has been dealt with extensively in the previous paragraph, and will not be further elaborated here.

If, for now, the premises of the critical realists are accepted, it can be understood that they assert that relationships between prognostic factors and outcomes are always probabilistic, and that these relationships can be known to human beings, yet only partially and imperfectly. Statements about the nature of reality do not take the form of simple cause and effect relationships (Everitt and Hardiker 1996). So, a Critical Realist would not argue that, in the case of A, when we apply method B, result C will occur. Instead they would argue that in the case of A, when we apply method B, result C is more likely to occur than when method X, Y or Z is applied. Thus, there are no law-like generalisations but only more or less general statements (Smith 2002). It is also acknowledged that these general statements tend to be situational and that we should not strive to find relationships between mechanisms and outcomes, but context – mechanism – outcome combinations (Frost 2002; Smith 2002).

Because social work and child protection inherently deal with situational, contextual issues, it might then be asked whether any evidence that has been produced in one place can be used in the other. If this is not the case, then the problem might arise that all evidence should be gathered or at least validated locally. That would then require lots of resources and, if an experimental approach is chosen, the problem could also arise that there are not enough subjects in the population to reach the required statistical accuracy. Thus, the feasibility of gaining knowledge about context – mechanism – outcome combinations can be questioned.

Furthermore, as far as the aforementioned kind of research strives to find probabilities in the strictly statistical sense of the word, it has to deal with the same problems that any quantitative approach is confronted with. The problems of standardisation, operationalisation, and numerous confounding variables pose a serious threat to validity. The difference between advocates of an experimentalist approach and those using a pragmatic, probabilistic, although quantitative approach to social research seems to be that the latter are better aware of the relativity of the outcomes of quantitative research and refuse to put one kind of research above the other.

Others propose a Pluralist methodological approach in which induction and deduction are both considered to make a significant contribution to professional knowledge development as part of a recursive process in which deduction answers questions and induction generates them (Taylor and White 2001). As a consequence, a methodologically inclusive stance is taken. Both qualitative and quantitative approaches are accepted and seen as complementary, depending on what question is

to be answered and as long as research has been conducted in a rigorous manner (Ainsworth and Hansen 2002; Gibbs and Gambrill 2002; Tierney 2002). It is argued that if 'what works' were the only question to be asked, quantitative experimental approaches would likely be a good method to arrive at an answer. Nonetheless, because it is also worth knowing what element of a certain intervention works, for whom it works, how it works, why it does sometimes not work, how service users experience a certain intervention, etc., qualitative approaches can not be done without (Humphries 2003). As has been argued earlier on in this paper, in the case of social work quantification entails reductionism. In order not to lose any other valuable information in the process, qualitative methods are needed. Moreover, qualitative, heuristic research is deemed indispensible because it generates the hypotheses that are tested in clinical trials.

The idiosyncratic application of nomothetic data that are gathered by means of RCTs is not thought to be straightforward because complex interactions, partial understanding and moral judgements are involved (Humphries 2003). For instance, the nature of the problem of the client is taken for granted in EBP, whereas in practice it is not clear-cut (Glicken 2004; Taylor and White 2001). So, if we take a look at the answering of clinical questions, which lies at the heart of EBP, the very first part of the clinical reasoning process is considered to be very difficult to answer. We don't always know what is the matter. Deciding what is going on is mostly a judgement based on incomplete and fallible information.

In addition, it is asserted that the problem of a specific service user in a real-life case is usually different from the problems of those who have participated in the trials that have provided the evidence to be applied. Thus, an intervention that has been proven to work is also going to work with a specific client. In the specific circumstances of the case, there might be an element that is a previously unnoticed contra-indication, because it might make the intervention useless, or even harmful. Furthermore, the specifics of the case will require the worker to pass judgement about the question whether a specific client fits the profile of somebody who could be helped with an intervention that has been 'proven' to work.

Methodological pluralism, although it is philosophically contested, is supported by many scholars and practitioners (see e.g. Ainsworth and Hansen 2002; Butler 2003; Humphries 2003; Taylor and White 2001; Tierney 2002) who believe it will render social work methodologies that can more appropriately capture social work reality.

It is not only the paradigm from within which scholars operate that shapes their arguments for or against EBP, it is also their interpretation of what EBP entails. Since EBP is not a commonly accepted and well-demarked concept, there can be no sensible debate about EBP as such. Discussions will, in the best case, be about interpretations of EBP.

It can be concluded that two of such interpretations have arisen over time: An experimentalist version that employs a hierarchy of evidence, and a pragmatic approach that is more methodologically inclusive. Looking back at Sackett et al.'s (1997) definition, it could be argued that the experimental approach limits itself to the question of what constitutes "best evidence" about "what works". The more comprehensive, pragmatic approach also asks itself how we can "explicitly,

conscientiously and judiciously use" that evidence in practical decision-making. Those who advocate the latter angle define EBP in a more integrated way and take into account that application of research is not straightforward, and that the value of experimental research is relative, not more objective than other kinds of evidence, and cannot be placed in a hierarchical relationship to other forms of research.

The above provides arguments for the problematic nature of strictly experimental EBP in the field of social work. No consensus seems to exist about what counts as good research and best evidence. This shows that social work is not in a state of 'normal science' in the Kuhnian sense of the word, and that the expertise that is based on various kinds of knowledge is contested.

Ethical Issues in EBP

Except theoretical arguments, ethics are often referred to as argument to either embrace or dismiss experimental EBP. The ethical issues involved in EBP centre around four themes: Harm, empowerment, professionalisation and cost-effectiveness. Many of the ethical arguments of the adherents of EBP lean on the presumed epistemological superiority of an experimentalist approach. As it has been shown that it does not make sense in social work research to put one kind of research above the other, these ethical arguments can not all be fully held. This will be discussed below.

First, Do No Harm

EBP-adherents argue that workers have a moral obligation to stop ineffective and counterproductive interventions (McNeece and Thyer 2004; Sheldon and Chilvers 2002; Tierney 2002) and that only practice that has been subjected to rigorous effectiveness research can claim to be ethical (Ainsworth and Hansen 2002). The expertise of the worker, which is often said to be based on whim and intuition, and the narratives of service users are not considered to be enough to base decisions upon (Zayas, Gonzales and Hanson 2003; Rosen 1994; Zeira and Rosen 2000; Gibbons 2001; Ainsworth and Hansen 2002; Newman and McNeish 2002).

That workers often do not base their practice on research evidence is warranted by an American empirical study that shows that one third of practice decisions are not underpinned by any kind of rationale, and that of those who were supported by a rationale, only 2% were based on any kind of empirical research. (Rosen 1994; Rosen et al. 1995). A Swedish study (Bergmark and Lundstrom 2002) has found that, although those who work in child and family services are a little more likely to read than others, generally only 10% of the respondents (n = 412) had ever heard of the three most famous scientific social work journals in their own language. This could make social workers susceptible to the transfer of untested, commonsense, "lay habits" (Rosen 2003) to situations in which well-founded professional judgement is required.

On the other hand it is argued that not being able to articulate a rationale does not equate to being atheoretical (Fisher 2002), although it might be an indicator of how much thought has been given to the methods employed and of whether reference is being made to external evidence. Having a professional theoretical basis is considered to be important by most. However, a lack of formal knowledge could be complemented by a heightened degree of reflexivity on behalf of the worker (Sheppard 1998; Taylor and White 2001). Furthermore, theory can also be seen as an

obstacle to professional judgement, as employing a particular theory as a panacea might lead to problems because issues are framed in terms of the theory. This could then lead to a search for confirming evidence and ignoring disconfirming evidence for the sake of maintaining the theory. Thus, workers can be turned into mindless machines who do what theory prescribes, but are not sufficiently self-critical or reflexive.

When workers do read relevant literature, this does not automatically imply that they will also use it when presented with a case in a practice environment. Workers should also learn how to identify, read and assess relevant research literature (Tierney 2002). An empirical multi-methods study in Australia found that social work was the profession that rated lowest on a comparative study of various professions knowledge of the available evidence-base (Murphy and McDonald 2004).

In other words, it is asserted that the "first, do no harm" principle should be hailed (Oakley 2002) and Social Workers should strive to know whether what they are doing is effective in terms of reaching the goals that were set at the beginning of the intervention. Once they have rigorously researched whether what they are doing actually works, practitioners are more accountable for the interventions they propose, and clients can be given the opportunity to make a well-informed choice.

However, as has been shown in the paragraph on the paradigmatic controversies surrounding EBP, what can be considered 'rigorous research' is a very contentious issue. The views on what constitutes 'best evidence' or 'rigorous research' cover the entire spectrum, ranging from an approach in which a clear preference is given to Randomised Controlled Trials (RCTs) to one in which all nomothetic data are regarded as useless evidence that cannot be applied to idiosyncratic cases.

In addition, there is no convincing evidence that social work practice that is conducted in a predominantly experimentalist, medical-biological, evidence-based manner is actually more effective than other forms of practice, whereas there are examples of the implementation of empirically supported approaches that do not necessarily result in better outcomes for service users (Richey and Roffman 1999). Furthermore, this question is not beyond the paradigmatic debate discussed before, as the methods that are to be used to assess the effectiveness of EBP will be either experimentalist or pragmatic, and will thus be disputed by one of both schools. From this it can be concluded that ethical arguments that are based on the presumption that experimental evidence-based social work practice is more effective than other, more inclusive approaches to evidence, cannot be held.

EBP as Empowerment or Oppression?

Another moral argument that is raised is that social work runs the risk of being deprived of its potential to contribute to the 'radical, emancipatory and transformative ideal'. It is deemed dangerous that governments are involved in social work research as subsidisers, because it could transform social work into a form of 'welfare practice' (Butler 2003). Problems are said to be individualised; malfunctioning of context or social systems is almost ruled out as a cause for behaviour (Humphries 2003; Wachholz and Mullaly 1997).

Those who regard experimentalist EBP as a technicist approach argue that its supporters presuppose that problems are neutrally defined whereas they are socially and politically defined. In that way EBP leads to a situation in which clients' opinions cannot play a role and thus undermines partnership and negotiation (Frost 2002). In other words, EBP is seen by some as an attempt to depoliticise social work because the goals that are to be attained are imposed on the client by politicians, managers and professionals on the basis of 'evidence' about 'what works'. Depoliticising EBP would amount to an infringement of social work ethics, because the purpose of social work should be to take sides with the powerless and not to impose politically or managerially endorsed outcomes on unsuspecting clients.

This argument can be sustained when it comes to determining the structural causes of problems that are experienced by service users, and the justification of the goals that will be set. From the point of view of structural social work, which embraces the idea that research should be geared towards empowerment of the oppressed, the goal of societal, rather than individual change should be the starting point for research (Wachholz and Mullaly 1997). It can be argued that it does not hold true that methodologies like RCTs *in themselves* serve certain stakeholders. RCTs, as all other methods, are merely means of knowing (Humphries 2003). Nevertheless, any evaluation method can be misused for serving political interests because the process of generating evidence about practice is full of value judgements (Everitt and Hardiker 1996). Goal setting is a moral activity in the sense that one political group may have benefit of attaining that goal, whereas another group may find it disadvantageous. Thus, any goal can be controversial because conflicting and irreconcilable interests are involved (Hermans 2005).

Because the concept of empowerment embraces both the facilitation of personal change and the exertion of influence in the socio-political arena by means of collective social action (Graham 2004), it need not be so that a choice has to be made between decrying societal causes of undesired behaviour, and an approach aimed at correcting such behaviour. Addressing social injustice as a cause of problems does not dismiss the worker from their obligation to also fight the symptoms and fight them in the best way they can. Analogically, in the medical world a physician would not withhold pain relief medication from a patient suffering of a tumour with the argument that it is the tumour that should be fought and not the pain.

Thus, interventions that are geared towards the adjustment of undesired behaviour could be argued to be of a different order than agitation against social injustice. This is to say that it would be ethically defendable to choose to confront social injustice *and* modify behaviour, but that it would be morally wrong to only choose a behavioural approach in which structural causes of problems are ignored. To again draw the parallel with the world of medicine: It would be considered totally irresponsible if a doctor were to cure the pain, without trying to annihilate the tumour. Moreover, it is asserted that even advocacy social work itself can be evidence-based. Highly political and controversial goals can be formulated, operationalised and evaluated (Ainsworth and Hansen 2002; McNeece and Thyer 2004). In reality however, one could ask oneself how likely a government is to fund research that will probably yield criticism of that same government.

Other arguments in favour of the democratising nature of EBP are that it is a prerequisite for informed negotiation between practitioner and client because EBP discards 'unfounded' authoritative knowledge, and opinions are only respected when empirically supported. That way, service users are enabled to critically assess the interventions proposed to them and are made less dependent of professionals' judgement (Newman and McNeish 2002; Tierney 2002). In other words: service users can be empowered by EBP.

On the other hand it could be argued that, because of the complexity of the interpretation of (especially statistical approaches) to social research, clients will become more dependent of workers, and workers will become more dependent of scientists and other experts. As has been shown above, researchers and practitioners have great difficulties to achieve consensus about what is to count as evidence or best evidence. It could then be asked how realistic it is to think that a lay person could interpret these data, and to think that the practitioner does not have their own criteria for what counts as rigorous research. Because research outcomes are always ambiguous and contestable, and can be interpreted in different, contradictory ways, one kind of authority is replaced with another.

According to MacDonald (1999), "opponents of EBP [who try] to leave the status of knowledge as purely a matter of individual preference, rather than an important technical and ethical issue, place the least powerful groups in society at risk of a double standard of operating by those more securely placed". A similar statement has been made by Oakley (2000) on the use of experiments in the social field.

In fact, these experimentalists argue that a Relativist approach which regards all evidence as equal when it comes to establishing 'what works', amounts to elitism. If every kind of research is as good as any other, decisions about what counts as a fact and what does not, becomes an issue of power. Only the most powerful will then be able to define reality. Thus it is argued that the only alternative to social work that is based on evidence is practice that is based on authority. (Gambrill 1999; O'Connor Murphy and McDonald 2004). However, the issue that goal-setting, construction of parameters, and choices regarding statistical analysis are authoritative value judgments as much as the designs of any other form of research is not convincingly dealt with by these authors.

Thus it becomes clear that the ideology of social work can be, and *is* contested. This given makes a consensus about goals between relevant actors in the professional field highly unlikely, and poses serious problems for the implementation of experimentalist evidence-based practice that is based on standardised goals and standardised measures.

EBP and the Professionalisation of Social Work

A further political issue is that of the relationship between EBP and the interests of the professional group. EBP is sometimes considered to be the lifeline of social work, as other approaches have not been able to bring professional status and privileges (Ainsworth and Hansen 2002; Humphries 2003). On that basis, it is argued that if social work does not embrace EBP, the idea that what social workers do can be done without, is likely to gain popularity (Murphy and McDonald, 2004; Witkin and

Harrison 2001). Because of its appreciation of rigorous research as the basis for professional decision-making, it is asserted that EBP can protect staff against negative images of the public (Tierney 2002; Taylor and White 2002). It should be countered here that, first of all, the status that EBP could give to social work can only be a secondary argument for adoption of that concept. The rise of social work on the professional ladder (if any) as a result of an evidence-based approach to practice will be a side effect of a possible increase of effectiveness. If EBP would be shown to bring no improvement to clients' situations, it would not be morally correct to maintain it anyway for the sake of the status of the profession.

Also, as has been argued previously in this chapter, there are arguments that quantification and experimentation are not any more rigorous than other kinds of research.

Moreover, whether social work is led by either experimentalist evidence-based or other, more inclusive methodological principles will not likely be an issue in the shaping of the public's opinion. EBP does not pretend to bring certainty, but it tries to increase probabilities (Newman and McNeish 2002). Therefore, even if it does prove to be a superior approach, evidence-based social work cannot rule out dramas like child deaths or any other incidents that are expected to have a more important influence on the general attitude towards social work.

EBP and Cost-Effectiveness

As has been said, one aspect of EBP is that it can also strive to provide value for money (Cheetham et al. 1992). It is argued that the market imperative is the dominant ideology and that the 'what works' movement strives for most effect with least as possible costs (Humphries 2003; Van Yperen 2005). By rigorously testing whether what social work agencies are paid to do actually works, and by optimising interventions, tax money can be saved. Some perceive this aspect of EBP as threatening, arguing that it is an instrument of managerial control (Webb 2001) and can serve as an instrument for reducing costs. On the other hand, effectiveness research can also be used as an argument by social work agencies to demand more money in order to be able to provide a more effective, yet more expensive intervention (Gibbons 2001; Gibbs and Gambrill 2002; Sackett et al. 1997).

As it has been asserted above, any method can be used for political purposes. Hence, a distinction can be made between managerially driven EBP that is focused on performance and efficiency on the one hand, and "evaluation for good practice" that is geared towards improving services on the other hand (Everitt and Hardiker 1996; Webb 2002).

The problem with the issue of controlling the costs, is that the costs of an intervention are often not juxtaposed to the effectiveness of the intervention. As a result, cost-effectiveness is defined as the relationship between costs and the *amount* of service provision, instead of the *effectiveness* of the services provided (Cheetham et al. 1992).

Whether EBP is judged as a potentially empowering concept or as an attempt by politicians and managers to turn social work into a tool of normalization and behavioural control depends on what research is geared toward, as well as the degree to which various forms of research are accepted as evidence.

It can again be concluded that ideology is not only contested when it comes to the issue of emancipating or controlling the client. It is also contested when it comes to emancipating the profession or keeping it under managerial control. Again, this poses difficulties to an experimentalist approach that aims to standardise goals and parameters.

Conclusions

Risk-aversion has become a key feature of social work in general, and specifically in the field of child protection, the call for evidence-based practice has poured over from the medical field into social work.

Because of the nature of social work, a strictly experimental approach to EBP in social work and child protection is impractical. However, this does not mean that experimentation is useless. The results of experiments should not be regarded as the rock bottom of knowledge, as their results are as much socially constructed as qualitative findings are. Social work is likely to have better chances of achieving an evidence base by means of a more methodologically inclusive, pragmatic variant of EBP.

It is clarified that there are also forces other than the quest for effectiveness that drive EBP. Although the ethical issue of harm reduction is closely related to epistemological stances, the issues of emancipation, professionalisation, and cost-effectiveness that are outlined in this paper, show that both the ideology of social work and the expertise of social workers are contested. As disagreement about ideology and expertise complicate the process of goal setting and the identification of what is to count as relevant knowledge, it can be concluded that the evidence-base of social work should be made up of research that generates multiple forms of knowledge and appeals to various ideologies.

References

Ainsworth, F., and Hansen, P. (2002). Evidence-based Social Work Practice: A Reachable Goal? *Social work and Social Sciences Review, 10*(2), pp. 35-48.

Baeten, P., and Willems, J. (2004). *De Maat van Kindermishandeling; Richtlijnen voor het Handelen van Beroepskrachten*. Amsterdam: Uitgeverij SWP.

Bergmark, A., and Lundstrom, T. (2002). Knowledge and attitudes to knowledge of social services personnel. *Social Work Education, 21*(3), pp. 359-373.

Buckley, H. (2003). *Child Protection Work; Beyond the Rhetoric*. Philadelphia: Jessica Kingsley Publishers.

Burnard, P., and Hannigan, B. (2000). Qualitative and Quantitative approaches in mental health nursing: moving the debate forward. *Journal of Psychiatric and Mental Health Nursing*(7), pp. 1-6.

Butler, I. (2003). Doing good research and doing it well: ethical awareness and the production of social work research. *Social Work Education, 22*(1), pp. 19-30.

Corby, B. (1993). *Child Abuse:Towards a Knowledge Base*. Buckingham: Open University Press.

Everitt, A., and Hardiker, P. (1996). *Evaluating for Good Practice*. London: MacMillan.

Fisher, M. (2002). The Social Care Institute for Excellence: the role of a national institute in developing knowledge and practice in social care. *Social work and Social Sciences Review, 10*(2), pp. 6-31.
Frost, N. (2002). A problematic relationship? Evidence and practice in the workplace. *Social Work and Social Sciences Review, 10*(1), pp. 38-50.
Gambrill, E. D. (1999). Evidence-Based Practice: An Alternative to Authority-Based Practice. *Families in Society: The Journal of Contemporary Human Services, 80*(4), pp. 341-350.
Gambrill, E. D. (2001). Social Work: An Authority-Based Profession. *Research on Social Work Practice, 11*(2), pp. 166-175.
Gambrill, E. D. (2003). Evidence-Based Practice: Sea Change or the Emperor's New Clothes? *Journal of Social Work Education, 39*(1), pp. 3-23.
Garretsen, H., Rodenburg, G., and Bongers, I. (2003). Evidence-based werken in de welzijnssector. *Sociale Interventie, 12*(4), pp. 30-35.
Geddes, J. (2000). Evidence-Based Practice in Mental Health. In L. Trinder and S. Reynolds (Eds.), *Evidence-Based Practice: A critical appraisal*. Oxford: Blackwell Science.
Gibbons, J. (2001). Effective practice: Social Work's Long History of concern about outcomes. *Australian Social Work, 54*(3), pp. 3-13.
Gibbs, L., and Gambrill, E. D. (2002). Evidence-Based Practice: Counterarguments to Objections. *Research on Social Work Practice, 12*(3), pp. 452-476.
Gilgun, J. F. (2005). Four Cornerstones of Evidence-Based Practice in Social Work. *Research on Social Work Practice, 15*(1), pp. 52-61.
Gira, E. C., Kessler, M. L., and Poertner, J. (2004). Influencing Social Workers to Use Research Evidence in Practice: Lessons From Medicine and the Allied Health Professions. *Research on Social Work Practice, 14*(2), pp. 68-79.
Glicken, M. D. (2004). *Improving the Effectiveness of the Helping Professions; an Evidence-Based Approach to Practice*. London: Sage Publications.
Graham, M. (2004). Empowerment revisited - social work, resistance and agency in black communities. *European Journal of Social Work, 7*(1), pp. 43-56.
Greenhalgh, T. (1999). Narrative based medicine: Narrative based medicine in an evidence based world. *British Medical Journal*(318), pp. 323-325.
Guba, E. G., and Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin and Y. S. Lincoln (Eds.), *Handbook of Qualitative Research*. Thousand Oaks: Sage.
Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, D. J., and Cook, R. J. (1995). Users' Guides to the Medical Literature. IX. A Method for Grading Health Care Recommendations. Evidence-Based Working Group. *JAMA, 274*(22), pp. 1800-1804.
Hermans, K. (2005). Evidence-Based Practice in het Maatschappelijk Werk. *Sociale Interventie*(3), 5-15.
Humphries, B. (2003). What else counts as evidence in evidence-based social work? *Social Work Education, 22*(1), pp. 81-91.
Jones, J., and Gupta, A. (1998). The Context of Decision Making in Cases of Child Neglect. *Child Abuse Review, 7*, pp. 97-110.
Kelly, N., and Milner, J. (1996). Child Protection Decision-Making. *Child Abuse Review, 5*, pp. 91-102.
MacDonald, G. (1999). Evidence-Based Social Care: Wheels off the Runway? *Public Money and Management*(January-March), pp. 25-32.

MacDonald, G., and Sheldon, B. (1998). Changing One's Mind: The Final Frontier? *Issues in Social Work Education, 18*(1), pp. 3-25.
Massé, R. (2000). Qualitative and Quantitative Analyses of Psychological Distress: Methodological Complementarity and Ontological Incommensurability. *Qualitative Health Research, 10*(3), pp. 411-423.
McNeece, C. A., and Thyer, B. A. (2004) Evidence-Based Practice and Social Work, *Journal of Evidence-Based Social Work,* vol. 1, no. 1, pp. 7-25.
Mead, G. H. (1934) *Mind, Self and Society*. Chicago: University of Chicago Press.
Melief, W. (2003) Evidence-Based Practice in Sociale Interventies, *Sociale Interventie*. vol. 4, pp. 17-29.
Munro, E. (2004) A simpler way to understand the results of risk assessment instruments, *Children and Youth Services Review*, vol. 26, pp. 873-883.
Murphy, A., and McDonald, J. (2004) Power, Status and Marginalisation: Rural Social Workers and Evidence-Based Practice in Multidisciplinary teams, *Australian Social Work,* vol. 57, no. 2, pp. 127-136.
Newman, T., and McNeish, D. (2002) Promoting Evidence-based practice in a child care charity: The Barnardo's experience, *Social work and Social Sciences Review*, vol. 10, no.1, pp. 51-62.
Oakley, A. (2000) *Experiments in Knowing: Gender and Method in the Social Sciences*. Oxford: Polity Press.
Richey, C. A., and Roffman, R. A. (1999) On the Sidelines of Guidelines: Further Thoughts on the Fit Between Clinical Guidelines and Social Work Practice, *Research on Social Work Practice*, vol. 9, no.3, pp. 311-321.
Rosen, A. (2003) Evidence-based social work practice: challenges and promise. *Social Work Research*, vol. 27, no.4, pp. 197-207.
Rzepnicki, T. L., and Johnson, P. R. (2005) Examining decision errors in child protection: A new application of root cause analysis, *Children and Youth Services Review*, vol. 27, pp. 393-407.
Sackett, D. L., Rosenberg, W., Gray, J. A. M., Haynes, R. B., and Richardson, W. S. (1996) Evidence-Based Medicine; What it is and what it isn't, *British Medical Journal*, vol. 312, pp. 71-72.
Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., and Haynes, R. B. (1997) *Evidence-Based Medicine; How to Practice and Teach EBM*. Edinburgh: Churchill Livingstone.
Sheldon, B., and Chilvers, R. (2000) *Evidence-Based Social Care: A Study of Prospects and Problems*. Lyme Regis: Russell House Publishing.
Sheldon, B., and Chilvers, R. (2002) An empirical study of the obstacles to evidence-based practice, *Social work and Social Sciences Review*, vol. 10, no.1 pp. 6-26.
Sheppard, M. (1998) Practice Validity, Reflexivity and knowledge for social work, *British Journal of Social Work*, vol. 28, pp. 763-781.
Smith, D. (2002) The limits of positivism revisited, *Social work and Social Sciences Review*, vol. 10, no.1, pp. 27-37.
Spratt, T. (2000) Decision Making by Senior Social Workers at Point of First Referral, *British Journal of Social Work*, vol. 30, pp. 597-618.
Taylor, C., and White, S. (2001) Knowledge, Truth and Reflexivity, *Journal of Social Work*, vol. 1, no.1, pp. 37-59.
Thyer, B. A. (1993) Social Work Theory and Practice Research: The Approach of Logical Positivism, *Social work and Social Sciences Review*, vol. 4, no.1, pp. 5-26.
Thyer, B. A. (2002) Popper, Positivism and Practice Research: A Response to Munro. *Journal of Social Work* Education, vol. 28, no.3, pp. 471-474.

Tierney, S. (2002) Reframing an evidence-based approach to practice. *Social work and Social Sciences Review*, vol. 10, no.2, pp. 49-62.
Trinder, L. (2000) A Critical Appraisal of Evidence-Based Practice. In L. Trinder and S. Reynolds (eds.), *Evidence-Based Practice: A Critical Appraisal*. Oxford: Blackwell Science.
Van Yperen, T. (2005) Van weten naar verbeteren, *SPH: Tijdschrift voor sociale en (ortho)pedagogische functies*, vol. 65, pp. 10-15.
Wachholz, S., and Mullaly, B. (1997) Human Caring; Toward a Research Model for Structural Social Work. *Canadian Social Work Review*, vol. 14, no.1, pp. 23-42.
Webb, S. (2001) Some Considerations on the Validity of Evidence-based Practice in Social Work, *British Journal of Social Work*, vol. 31, pp. 57-80.
Webb, S. (2002) Evidence-Based Practice and Decision Analysis in Social Work. *Journal of Social Work*, vol. 2, no.1, pp. 45-63.
Witkin, S. L. (1999) Questions, *Social work*, vol. 44, no.3, pp. 197-200.

Research Philosophy and its Influence on Research Process

T Mogotlhwane, F Khosrowshahi and J Underwood
School of Construction and Property Management

Introduction

This paper discusses the philosophical paradigm which forms the basis of the main research by the author. The philosophical stance is then linked to the main research, followed by the main research and methodology. The proposed research method to be used, the data generation methods and data analysis are also discussed.

The process of doing a PhD is confusing especially the first six months. A PhD candidate's is expected to do a literature review while at the same time learn and become knowledgeable on the appropriate research process. This is like rowing a boat, both literature review and learning the art of doing research needs to be balanced to avoid moving in a circle. However, this is not easy, especially to those who are not experienced researchers. The problem of confusing terminology as observed by Bell (1995: 5), Grix (2001: 1) and Crotty (2003: 1) makes research appear to be a difficult process.

Unfortunately, there is no other way of doing research especially at PhD level which can allow a researcher to avoid knowing the associated terminology. To illustrate this, Grix (2001: 26) compares understanding and knowing terminology of research to someone who wants to be a bricklayer. To a bricklayer, knowledge of difference as well as use among tools of the trade such as trowel and chisel is a must. Doing a PhD is of late regarded as learning how to do good research. With the advent of a globalised world, as observed by Gray (2004: 1), rapid changes in business and working conditions are continuously challenging organisations. To survive in this environment, organisations need to understand, hence conduct research to generate knowledge. The need to conduct research is probably higher than it has ever been due to rapid changes in technology and business practice in a globalised world.

Philosophical Paradigm

There are a basic set of beliefs that support human beings when they take action. Guba (1990: 17) refers to these set of beliefs as a paradigm. In research, ontology, epistemology and methodology are the basic paradigm that must be understood. Ontology and epistemology are very difficult concepts to understand. As noted by Grix (2001: 26), the two terms are greatly misunderstood. However, they form the research philosophy on which researchers formulate and base their enquiry. Saunders (2003: 82) claims that research philosophy is actually the first onion layer to be peeled in the research process. It must be understood first before the inner details can be unveiled. It is very difficult to draw the boundaries between ontology and epistemology.

Some of the examples of ontological stance that a researcher may take, according to Grix (2001: 26-27), are objectivisim and constructivisim. The objectivisim stance argues that reality exists on its own and is out there to be discovered. This position is

popular in pure science research. Constructivism on the other hand, according to Crotty (2003: 8), challenges the objectivisim stance and claim that knowledge is created as part of our interaction with reality. Crotty (2003: 5) puts objectivism and constructivisim under epistemology while Grix (2001: 26) treat the two as ontological stance. Ontology and epistemology are so closely related that they have a lot of common overlap. Due to the closeness of ontology and epistemology, Crotty (2003: 10) observed that authors have difficulty in separating them. This is confirmed by the fact that Saunders et al (2003: 82-112) does not mention the two words in his book and instead the author uses research philosophy.

Ontology

A detailed definition of ontology is that given by Grix (2001). Grix (2001) defines ontology as "the basic image of social reality upon which a theory is based. Ontology relates to the way in which an individual views the world" (Grix, 2001: 138). Grix goes further and states that ontology is connected to how a person finds out and generates knowledge. Oates (2006) refers to ontology as "assumptions about the nature of physical and social reality" (Oates 2006: 287). It is therefore the foundation of research. In this research, which looks at the impact of information and communications technology applications in performance of civil servants of Botswana, the ontological position adopted is that of constructivisim. This is because civil servants are people and they have their own sets of beliefs and values which cannot be discovered by carrying out experiments.

Epistemology

According to Grix (2001: 27), epistemology originates from two Greek words, namely episteme (which means knowledge) and logos (which means reason). Based on this, Crotty (2003) defines epistemology as "the theory of knowledge embedded in the theoretical perspective and thereby in the methodology" (Crotty 2003: 3). A similar definition is "one of the core branches of philosophy, is concerned with the theory of knowledge, especially with regard to its methods, validation" (Grix 2001: 27). According to Oates (2006), epistemology is "assumptions about the nature of knowledge and how it can be obtained" (Oates 2006: 287).

There are numerous epistemological positions as shown by Cresswell (2003: 4). Crotty (2003: 5) groups these positions together and calls them "theoretical perspective". A researcher will have his own ontological stance and each stance has its own preferred epistemological position. The two opposing epistemological positions are positivism and interpretivisim. Positivistic stance is the oldest mode of enquiry or paradigm. It uses pure science methods of investigation (scientific methods) and according to Oates (2006: 283) it assumes a non random, regular and orderly world, which can be investigated in an orderly manner.

Positivistic stance and its scientific method were developed to study the natural world, for example to carry out physics, chemistry or biological experiments. It is not ideal method for studying the social world of individuals, organisation and people as observed by Oates (2006: 288)

Interpretivism stance falls within the constructivism ontological position. This stance is taken by researchers who challenge the positivistic stance and "argue that social

world of business and management is far too complex to lend itself to theorising by definite 'law' in the same was as the physical science" Saunders et al. (2003: 84).

Reasons for doing PhD

There are a number of reasons that can lead to a researcher to undertake PhD. Personal experience, contribution to knowledge, solving a problem are some of the reasons as observed by Grix (2001: 57) and Oates (2006: 16-21) . For PhD one of the reasons must be "an original contribution to knowledge "(Phillips and Pugh 1994: 63).

For this research, the maim reasons for undertaking PhD are to:

- Contribute to knowledge
- Find evidence to inform practice

The second reason has been influenced by the author's experience as well as concern for poor performance among the civil servants of Botswana. The government of Botswana has taken a number of initiatives to address the problem, but none seems to bear fruits. Botswana has also invested heavily in the acquisition and development of information and communications technology infrastructure. The author intends to study what could be the possible reasons for such investments not to be yielding desired results of improved service delivery. According to Oates (2006: 18), sometimes managers acquire information systems like enterprise resource planning (EPR) just because it is fashionable to do so.

Broad Research Area

Introduction

This study will determine the impact of information and communications technology applications in productivity in selected government departments of Botswana. It is intended to support Botswana's vision of becoming an international information and communications technology hub. It is not always the case that when an organisation adopt and implement information and communications technology, it will achieve the desired goals of productivity gains and better work practices. According to Scharge (1997: 178), the investments in information technology have been based on a 'lie' which is so perfect that it is easy to believe. He was referring to the promise that has been widely published, mainly by the information technology practitioners on use of information technology in organisations.

In Botswana within the last 10 years, the government has invested heavily in providing relevant information and communications technology infrastructure in all government departments. This is especially true in the selected government departments. Since the early 1990's there has been an outcry for lack of productivity in most of government departments. In another research by Thapisa and Jain (2000: 86), where they looked at perceptions about productivity in academic and public libraries in Botswana, lack of productivity also came out. Botswana government took a number of initiatives to address lack of productivity in the civil service such as Work Improvement Teams (WITS), setting up Botswana National Productivity Centre, implementing performance based reward system, developing good information and communications technology infrastructure in government

departments, etc. however, the problem still exists today, despite relatively good information and communications technology infrastructure such as email, internet and telephone availability in government departments. According to Little and Bose (2004b), Botswana "is already an active participant in the global information society" (Little and Bose 2004b: 14).

There has not been any detailed study to find out why information and communications technology introduction in government has not had a positive impact in productivity. The introduction of information and communications technology in government departments has a history of being user driven where the core of it was mainly to automate the manual tasks. This research looks at issues of adaptation, strategy and acceptance of information and communications technology in central government. How these issues are addressed can determine the success or failure of information and communications technology. To address this problem, the following research question, hypothesis, aim and objectives have been developed.

Research Questions

The main question that this research tries to answer is:
Why information and communications technology applications are not having positive impact in service delivery by government departments of Botswana?

Main Hypothesis

A proper Exploitation of Information Technology could lead to significant productivity gains in public projects in Botswana

Aim

Develop framework for information and communications technology exploitation in aid of improving productivity in public projects in Botswana

Objectives

The objectives of the study are to:
determine the limiting factors to full utilisation of information and communications technology in government department in Botswana
determine customer satisfaction levels on services offered using information and communications technology
determine user level of satisfaction with information and communications technology use, support.
develop a methodological instrument to help government departments evaluate impact of information and communications technology applications.

Statement of the Problem

There is a need to conduct comprehensive research as to why information and communications technology has not improved productivity in government departments of Botswana. Gates (1999: 16) stated that "It can help you with your problems, but those problems won't disappear just because you have a computer". He was emphasising that information and communications technology alone cannot lead to productivity gains. Previous research which focused on determining the relationship between information and communications technology investments, productivity gains and work practices were inconclusive. This is because it is not possible to isolate information and communications technology input; there are other

inputs that contribute to productivity in organisation Brynjolfsson (1993: 67). It will not be fair to put all the blame for lack of productivity gains in information technology alone. Government departments lack appropriate methodologies and suitable instruments for analyses to ensure that information and communications technology and work practices are leading to productivity gains. The research is in line with the Botswana Government priority in the National Development Plan 9 (NDP 9), which is "focused on high quality service delivery and improving organisational performance"

Literature Review

The problem of whether investments in information technology lead to any improvements in productivity has been researched and debated for a long time Oz (2005: 789). Unfortunately a lot of these studies did not produce conclusive results Sircar (2000: 69). The question of information technology investments and its impact on productivity has puzzled the research community for a long time. However, organisations continue to increase their investments in information technology, despite not having any evidence as it is leading to any positive returns. This has been referred to in literature as the "IT productivity paradox" (Oz 2005: 789) .

Overview of Productivity Debates in Botswana

The debate on level of productivity in Botswana has a long history. It can be traced as far back as the 1980's. At that time there were very few computers in government departments as compared to date. Venson (1986) observed a number of factors that led to low productivity in Botswana and among them computerisation is not mentioned, because at that time it was not a tool of the trade. However, when looking at level of productivity in library services in Botswana, Thapisa and Jain (2000: 92) included lack of technology as one of the factors that contributed to the problem.

The debate intensified at the beginning of the 1990's, the president of the time Masire (1991), referred to the lack of productivity as "culture of laxity". It was through such sentiments that the government set up Botswana National Productivity Centre (BNPC) in 1995 to try and instil sense of desire to perform better among the nation.

There is a need to map the evolution of information technology in the civil service with the productivity debate in order to find how the two overlap. Computerisation in central government was introduced gradually from around the middle of the 1990's. The ICT infrastructure has been gradually improving over time. This was also observed by Duncombe and Heeks (2001: 1) who wrote that Botswana had "relatively well-developed ICT infrastructure, and its favourable policy environment". This is a clear indication that government has the drive to move into the information society of the 21st century.

ICT Development in Botswana

The deployment of information technology in central government up to now is fairly developed. A wide area network called Government Data Network (GDN) connects all government departments, which also provide high-speed access to the internet as observed by Little and Bose (2004a). The development of this infrastructure has been associated with huge investments. According to Little and Bose (2004a), investments in ICT in the last three development plans is as shown in Table 1.

NDP No.	Duration	Amount in Pula (P)	% of Government Budget
7	1989-1996	19.3M	.2
8	1996-2003	600M	2.7
9	2003-2010	1.3BN	3.7

Table 1. ICT Expenditure from NDP 7 to 9. Source: (Little and Bose, 2004a: B3)

It is now clear that the productivity debate started well before deployment of information and communications technology. The two are out of phase; however the expectation was that as information and communications technology diffusion increase, there will be a corresponding improvement in public service performance. Unfortunately this is not the case. Up to now there is sill debate on the level of productivity, which is not at a satisfactory level. Venson (2005) when speaking in a public meeting indicated that civil servants who do not perform to their best were denying the public service provided by government. Rammidi (2005) expressed similar sentiments and even suggested that members of the public must be allowed to assess performance of civil servants so as to detect non-performers. Maine (2005) in launching Ombudsman annual report stated, "even in the days of email, there is still delay in attending to official mail".

The problem of level of productivity in Botswana is seen by the public through comments raised in public forums e.g. Kgotla meetings, politicians in their debates in parliament (Sebetlela 2005), the business Community (Ibrahim 2005) and the academic community (Thapisa and Jain 2000). This shows that the problem is genuine and ways must be found to deal with it.

Information Technology Specific Issues in Productivity in Botswana

It can be observed that as information and communications technology diffusion increased in public institutions, there is still concern that level of productivity is below expectations. The interesting thing now is that there are instances where information technology is also blamed for contributing to the problem. In an article of Botswana Daily News of 25th July 2005, customers of Botswana Building Society complained about inconvenience associated with using the society's ATM. The main problem was that ATM was not up to date with customer balances in their accounts.

Livestock Identification System (LIS) is a computerised information system that was introduced in the Department of Animal Health and Production (Ministry of Agriculture) in 2001. It was introduced as a response to the demand by the European Union that required having a system in place that will allow back tracking of all beef products from Botswana entering the EU market. The Botswana government had to respond to such a demand to protect its lucrative market of EU, especially that the country has not been able to meet the agreed quota of beef supply to the EU.

In response to the above, the government introduced LIS in 2001. This system has its history in Australia. However, there is a difference in pastoral farming system

between the two countries. In Australia livestock is reared in commercial farms with fairly developed infrastructure developed, e.g. electricity, telephones, etc. and this enables farmers to be able to monitor their stock. In Botswana, most of the farmers use communal grazing areas where animals mix with each other and individual farmers are not able to keep good records about their livestock. Infrastructure is also not developed in communal areas as well as in the few farms. The LIS system is presenting some problems as observed by Raborokgwe (2005) who argued that lack of staff to implement the system effectively was contributing to delaying farmers to sell their livestock.

Methodology

The research methodology framework is as shown in Figure 1

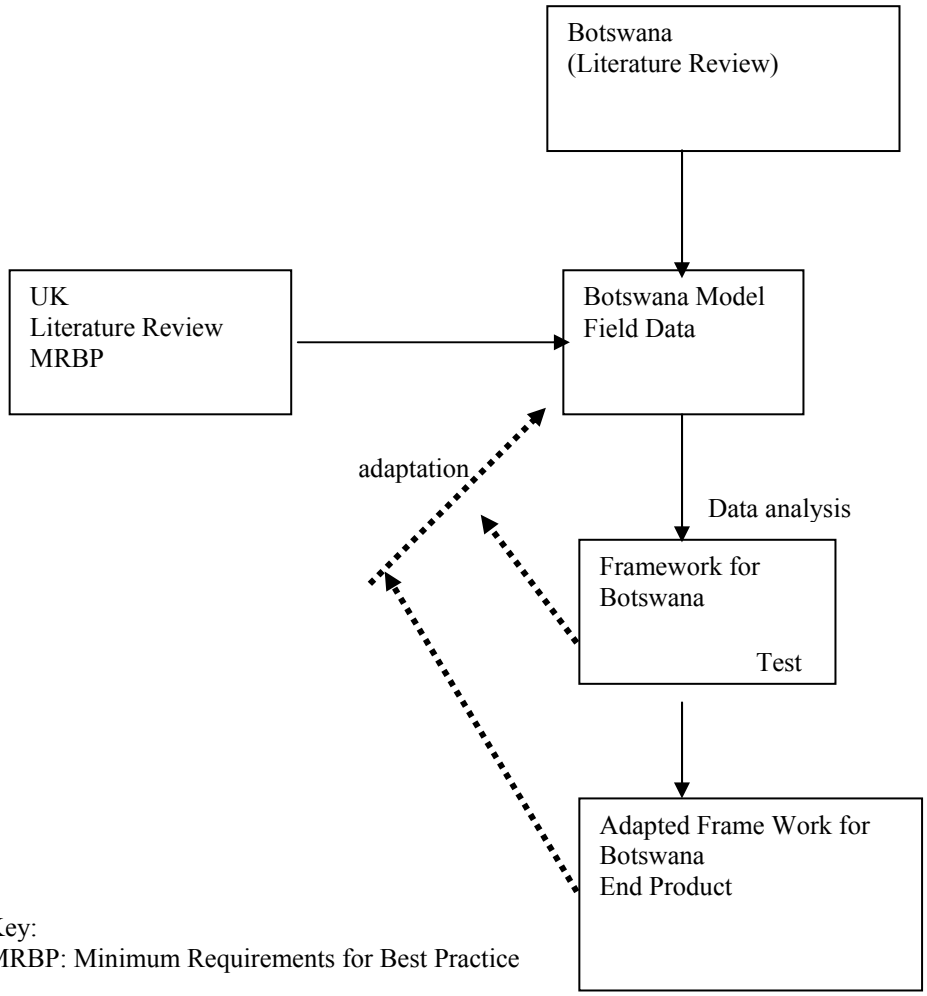


Figure 1: Research Methodology Framework

A thorough literature review on the problem of lack of productivity and information and communications technology developments will be undertaken. This will take a global view in order to find out if the problem has been researched previously. A detailed review of literature specific to Botswana will be undertaken during data collection.

In order to determine a benchmark against which to develop a model for Botswana, minimum best practice requirements for the UK government will be determined. UK

has a well-developed information and communications technology infrastructure with government services being made available online. As the two countries have different economic levels, the best practice determined will then be adapted to formulate a model for Botswana. The model for Botswana will then be tested and evaluated using Botswana data.

This research calls for the use of quantitative study approach to be able to evaluate previous records such as annual reports as well as current status to be able to test the hypothesis.

According to Crotty (2003: 2) any research has major four stages irrespective of type of research namely epistemology (what if means to know), theoretical perspective (assumptions about reality that influence research) methodology (type of plan to be used) and Method (type of instrument used to collect data). This is shown in Figure 2.

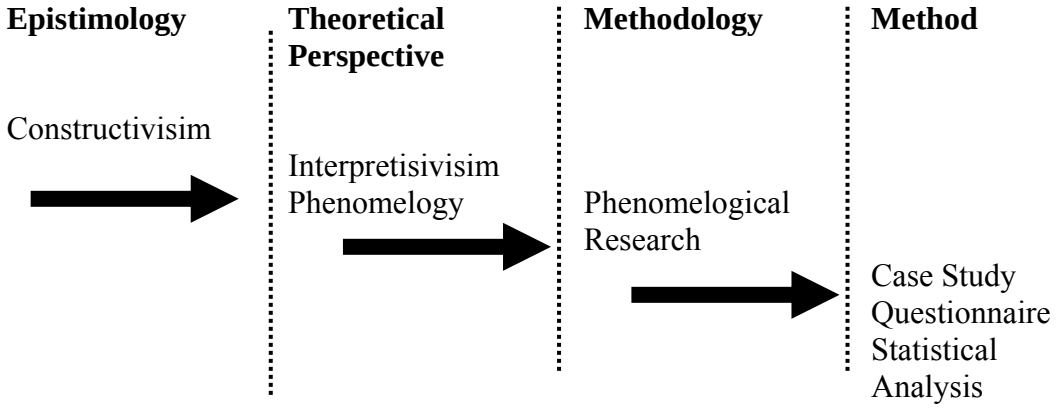


Figure 2: Stages of Conducting Research, Adapted from (Crotty, 2003: 5)

The research has been influenced by the researcher's personal experience as well as official and public concerns with the state of service delivery by some of central government urgencies. This research proposes that something can be constructed out of what is perceived as lack of satisfactory performance by employees. The research intends to adopt constructive epistemology to be able to reveal knowledge constructed through evaluation of interaction of people with information and communications technology. The theoretical perspective on which the research is based is the interpretisivisim as the researcher wants to base the theory in users, public and management experiences.

Phenomelogical approach is used to avoid the researcher using his personal perception as well as that obtained in literature to influence people's views. According to Gray (2004) "phenomenology insists that we must lay aside our prevailing understanding of phenomena and revisit our immediate experience of them in order that new meanings may emerge" (Gray 2004: 21). Knowledge will be interpreted through carrying out phenomelogical research using case study method. Questionnaire, structured interview, semi-structured interview, focus group as well as document analysis will be used to collect data and later analysed using statistical methods through the use of

SPSS for quantitative data. There will be a pilot study to test the method used as well as questionnaire to be used. The actual data collection will take place in Botswana

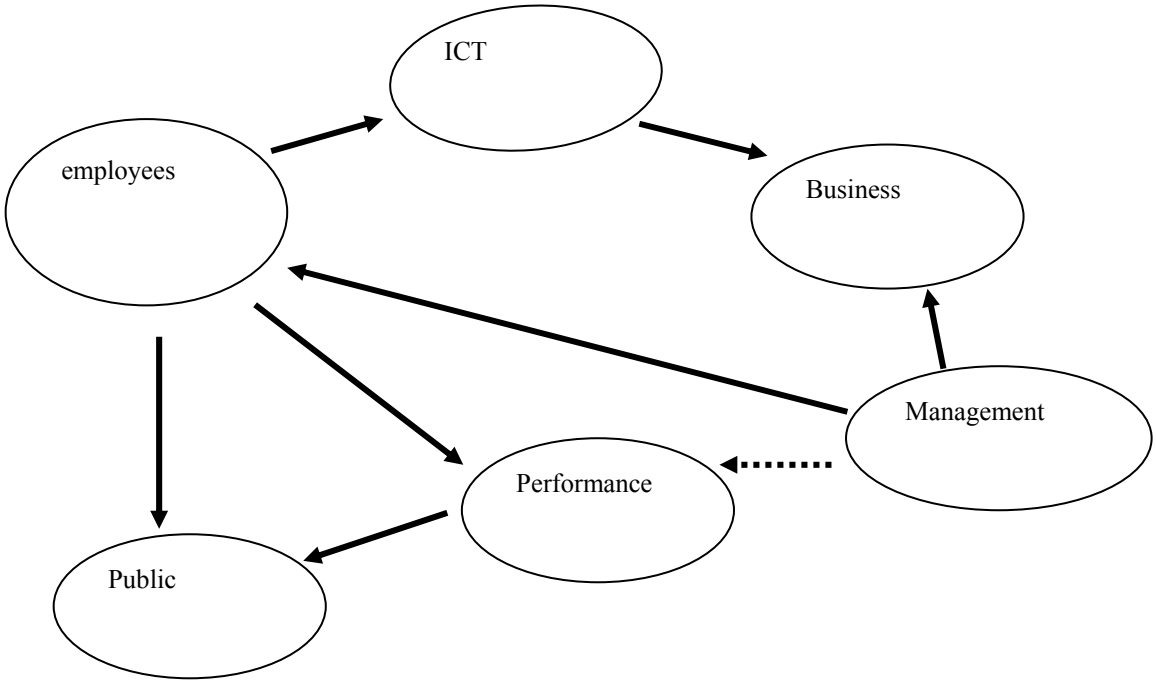


Figure 3: Research Conceptual Model

Conclusions

This paper has given a background on philosophical paradigm associated with the research to be conducted by the author. It also established the philosophical stance that the researcher takes. The most popular philosophical stance is that of positivisim which uses the scientific methods of investigation. This paper has demonstrated that it takes the opposing view to positivisim by using interpretivisim. It is rear in academic work for a researcher to state their philosophical position, but in looking at their research strategy and data generation methods, it can be established which philosophical paradigm is behind the research. The paper then links the interpretivisim stance to the main research and methodology. The focus for this research is on issues related to people. In trying to understand the source of the problem of low productivity in civil service of Botswana the issues concerning people is important as they are either the facilitators or consumers in the government machinery. It will also determine the level of satisfaction by the public on government services. For employees, their user satisfaction level and education on information and communications technology application and management will be covered. These areas have a link to performance as shown in Figure 3.

References

Bell, J., (1995) *Doing Your Research Project*. Buckingham: Open University Press

Brynjolfsson, E. E., (1993) The productivity paradox of information technology, *Communications of the ACM,* vol. 36, no. 12.

Cresswell, J. W., (2003) *Research Design: Qualitative, Quantitative and Mixed Methods Approaches.* 2nd Edition Thousand Oaks, CA: Sage.

Crotty, M., (2003) *The Foundation of Social Research: Meaning and Perspective.*London: Sage.
Duncombe, R. and Heeks, R., (2001) Information and Communication Technologies and Small Enterprise in Africa. Manchester, Institute for Development Policy and Management.
Gates, B., (1999) *The Road Ahead.* 3rd Edition. Harlow: Longman.
Gray, D. E., (2004) *Doing Research in The Real World.* London: Sage.
Grix, J., (2001) *Demystifying Postgraduate Research from MA to PhD.* Birmingham: University of Birmingham Press.
Gube, E. G., (1990) *The Paradigm Dialogue.* London: Sage.
Ibrahim, I., (2005) Ibrahim criticises PBRS, Botswana Daily News, Gaborone, Botswana.
Little, D. and Bose, A., (2004a) Maitlamo, Appendix B, Government e-Readiness. Gaborone, Department of Information Technology.
Little, D. and Bose, A., (2004b) Maitlamo, National E-Readiness Assessment, Volume 1 - e-Readiness, Executive Draft. *Botswna National ICT Policy.* Gaborone, Depatmetn of Information Technology.
Maine, L., (2005), Delays in attending to official mail worrisome, *Botswana Daily News*, Gaborone, Botswana.
Massire, Q. K. J. (1991) Address by president to 1st Meeting of 3rd session of the 6th Parliment of the Republic of Botswana. Gaborone, Botswana.
Oates, B. J. (2006) *Researching Information Systems and Computing.* London: Sage.
Oz, E. (2005) Information technology productivity: in search of a definite observation, *Information and Management,* vol. 42, no. 6**,** pp. 789-798.
Phillips, E. M. and Pugh, D. S., (1994) *How To Get a PhD: A Handbook for students and their supervisors.*3 rd Buckingham: Open University Press.
Raborokgwe, M., (2005), Cattle bolus proves a problem, *Botwana Daily News*, Gaborone
Rammidi, K., (2005), Involve public in assessing officers performance, *Botswana Daily News*, Gaborone, Botswana.
Saunders, M., Lewis, P. and Thornhill, A., (2003) *Research Methods for Business Students.* 3rd Edition Harlow: Prentice Hall
Schrage, M., (1997) The Real Problem with Computers. *Harvard Business Review,* vol. 75, no. 5.
Sebetlela, B., (2005), PMS bound to fail, *Botswana Daily News*, Gaborone, Botswana
Sircar, S., Turnbow, J. L. amd Bordoloi, B., (2000) A Framework for Assessing the Relationship Between Information Technology Investments and Firm Performance. *Journal of Management Information Systems,* vol. 16, no. 4, pp. 69-97.
Thapisa, A. P. N. and Jain, P., (2000) Bio-factor analysis of library productivity. *Library Management,* vol. 21, nos. 1/2, pp. 86-93.
Venson, P., (1986) *Productivity in the public service, fact of fallacy,* MSc Thesis, University of Central Michigan.
Venson, P., (2005), Report Poor Performers, *Botswana Daily News*, Gaborone, Botswana.

Is Your Degree Worth the Effort? Supply Of and Demand For Higher Educated Manpower in the UK

Michael Quigley

Management and Management Sciences Research Institute

Introduction

One of the UK government policies for higher education is for 50% of 18 to 30 year olds to have participated in undergraduate schemes by 2010. The Department for Education and Skills (DfES) cites that there is a necessity for this, for the UK economy to remain competitive with other "knowledge economies" (Clarke 2003). Many academics and economists believe that the 50% target is too high, for the British economy (Brown 2004). By using the system dynamics (SD) methodology a model of the interactions between employment and education can be built and analysed.

By building a system dynamics model a greater understanding of the relationships between tertiary education and employment can be obtained. The SD model's structure will highlight the feedback loops and non-linear relationships present, in the system. The model can be used to help in the understanding of how policy decisions impact on part or the whole of the system. However, SD is not used to forecast accurate numbers but to mimic the behaviour of the system when compared to historic data (where possible) and output information for future time periods. By testing and checking the model structure it is possible to validate the model and have confidence that the model output is both reliable and behaves in the same manner as the actual system would, when subject to similar variable changes.

The model runs from 1996 to 2020. From 1996 to 2004 the model output can be compared to past (actual) data, for key variables, and helps validate the reliability of the model's output.

The model's structure is built from existing knowledge and literature, using a number of tools including boundary charts, high-level mapping, influence diagrams and stock/flow diagrams. The data is also taken from many sources, much of which can be found on the Higher Educational Statistics Agency (HESA) website which produces annual reports titled Statistics of Education.

Because there has been great care and consideration undertaken to make sure that the terms of use are those by the current government, some wording can seem misleading or be open to interpretation. Where appropriate the definition of terms used has been stated in a small glossary at the end of this paper.

Background

The first section of this paper examines key points of the UK government's White Paper on higher education and how these are different from the current situation.

In January 2003 the Department for Education and Skills (DfES) in the UK government published a white paper entitled The Future of Higher Education (Clarke 2003). The paper identifies areas where it feels improvements can be made:

- The view of the then Secretary of State for Education and Skills, Charles Clarke, was to encourage people from less advantaged families to enter universities.

- A further proposal was aimed at giving universities "the freedoms and resources to compete on the world stage", by providing financial stability.

- The final aim of the government is "to make the system for supporting students fairer". It is their belief that the best way of doing this is by asking students to contribute to the costs involved in studying a degree.

Each of these additional points indicates that the cost of university will rise for its participants. This rise in cost will be modeled after enrolment in September 2006 when accurate data can be obtained; though early studies indicate that enrolment numbers will not decrease in the immediate short-term.

There are a number of targets the government sets universities. One such target is "to increase participation towards 50 per cent of those aged 18-30, mainly through two-year work-focused foundation degrees", whilst not increasing the overall number of students choosing to 'drop out' of the education system.

The government goes on to quote figures given in a recent report, by (Wilson 2001), "that 80 per cent of the 1.7 million new jobs which are expected to be created by the end of the decade [2010] will be in occupations which normally recruit those with higher education qualifications".This is due to the governmental perceived knowledge based economy (KBE) of the UK.

There are a number of critics who suggest that these numbers are not telling the whole story, most recently in the book The Mismanagement of Talent: Employability and Jobs in the Knowledge Economy By Brown and Hesketh, who as far as suggest that the "United Kingdom's economy is far from being that of a KBE" (Brown 2004).

There is not a system dynamics (SD) model that has been developed to model tertiary education exclusively, though SD models exists where this is included as a section of a much larger model, one example of this is Towards a Transition to a Knowledge Economy: How System Dynamics Is Helping Sarawak Plan Its Economic and Social Evolution (Dangerfield 2005). In this paper Dangerfield incorporates higher education into a much larger model, which has been built to help the State of Sarawak's 'State Planning Unit' better understand and manage the transition of their economy from the production based economy (p-economy) to a knowledge based economy (k-economy). Dangerfield's model, on the whole, does not examine the driving forces behind the demand for higher educated manpower or where these people are supplied from, but rather the type of graduates the government would like higher educational institutes to supply.

The tertiary education sector of the Sarawak" model splits students into two groups, arts and science students. This is particularly important in Sarawak as an increase in science students would be required to increase the "knowledge" in the economy. The separation in this model uses individual stock and flows for each group (arts or science) whereas the author here uses a series of subscripts.

There is also no modelling of the different levels within tertiary education in the Sarawak model, with undergraduates and postgraduates held in the same level and adult-learners (so important in the UK economy) not included either. It is often the case to find work carried out at this level, and without explicit aims in mind it would be rare in SD to model to a greater depth than this. The author's proposed model will be looking at much more detail, and will eventually include individual subject areas. Dangerfield's model should not be ignored as it gave insights into structural design and dynamic behaviour that is also seen the author's model.

Most other SD models that have been used in relation to higher educational institutes (HEI) centre on other aspects such as the emergence of new management problems (Kennedy 1999) but are not been applied to the structure of the system or flow of students through HEI.

As with system dynamics, there is little in-depth literature on higher educational institutes (HEI) and how they interrelate to the employment sector. The majority of the studies in this field concentrate literature in HEI, with little regard to the driving forces that supply the students.

By far the most popular methodological approach in the field of higher education is questionnaires and case studies, with the UK's Quality Assessment Agency spending £8 million annually on such studies (Baty 1999). The majority of these studies are into the effects a degree has on future employment.

Many of the case studies that do get published argue against the government's position on the economy. Of these papers many are found printed in Studies in Higher Education. Over- and Undereducation in the UK Graduate Labour Market [sic] (Alpin 1998) cites that any perceived gaps in the labour market's knowledge are filled by existing employees rather than graduates, who lack the experience. Indeed some economists have stated that "the attainment of qualifications threatens to run ahead of the economy's ability to absorb those qualifications" (Robinson 1995). This work looks at the characteristics of the employees in the job market and concludes that the employment market is already over-educated. Studies like this often rely on the Job-Completion Model (Thurow 1972). Thurow's model is a framework of questions comparing the actual level of qualification required in the UK compared to the given level of qualification. Alpin did find some areas of the labour market that required more qualified persons, using the standard occupation classification (SOC) used by the government, and also the modal measure for people employed already.

Most of the above studies do not refute the idea of attaining higher qualifications, as the majority of these studies rely on case studies. An in-depth analysis has shown that there can often be advantages to over-qualification, as they lead people to strive for more work, increase productivity and offer addition skills (Mason 1996).

There are other methodologies, most noticeably econometrics, that model aspects of higher educated manpower or model individual subject areas in the tertiary education system. Much information and understanding can be obtained from these models, though a lack of system structure is usually evident, with such methodologies. SD models have a logical and clear structure, which is usually developed from the model's objectives. By clearly defining the objectives for the research, and by employing the use of a high-level map, it is possible to create a boundary chart. The boundary chart can also be displayed in a table format that allows variables to be quickly categorised as either "endogenous", "exogenous" or "excluded". For this reason the boundary chart has been labeled as the "Three E's Table" (TET).

Model Building

When model building, it is appropriate to state the primary objective of the research which is "to study the supply and demand for higher educated manpower in the UK".

There is a great amount of information available related to the supply of and demand for higher educated manpower. This information is often obtained from UK government websites such as the Higher Education Statistics Agency (HESA). The annual reports they produce not only release information on the numbers of people participating in education but also the possible alternative routes that people can take. One example is the "high school leavers" who can choose to enter further education, unemployment or employment (in unskilled work). Absorbing, interpreting and expressing this information in an SD model requires a much more in-depth knowledge of the individual model parts than will initially be built – so that future model expansion is not constrained by past work. However, an extended model boundary and too great an emphasis on detail could make the model to difficult to build, display or understand. The latter problem is often referred to as the "black box" problem.

To help set model boundaries and identify key areas of the proposed model a high-level map was formulated (figure 1). This high-level map expresses the main inflows, outflows and interactions between education and employment. It is good practice to make SD models as holistic as possible and in order to do this all aspects of education, employment and the entities that affect them must be considered in respect of whether they are required in the model. The boundaries for this model were set to be those in the high-level map, though future work may require new boundaries, as the individual objectives will change. Greater detail on one subject area may be an example of this.

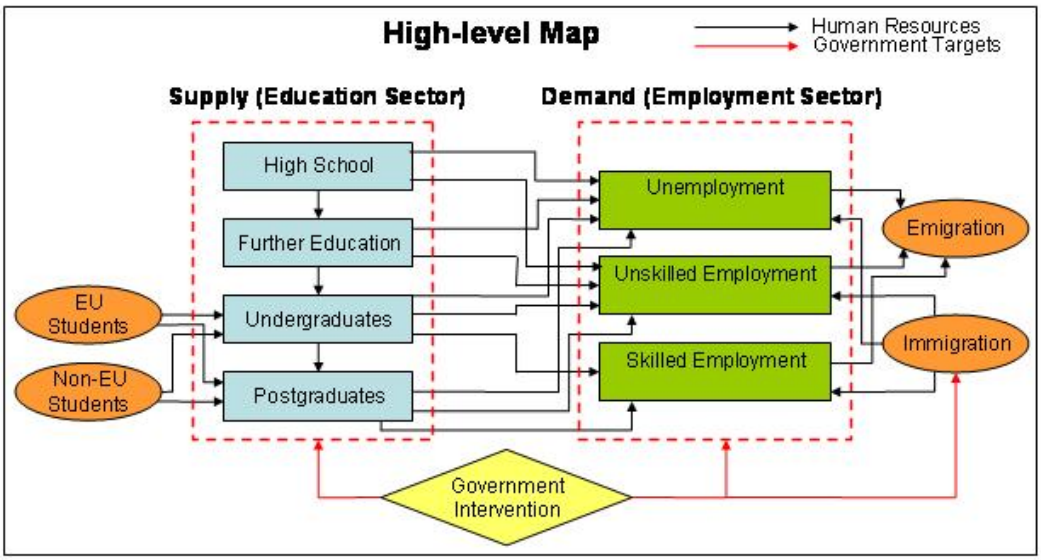


Figure 1: High-Level Map of Proposed Model

The high-level map has two defined sections: supply and demand. The supply of higher educated manpower is driven by the education sector and the demand from the employment sector. How these interrelate and are affected themselves is looked at in more detail as the model is defined.

A very simple SD model showing the flow of higher educated manpower, in the United Kingdom (UK), has been expressed in figure 2. The level of "Higher Educated Manpower" is increased as students graduate and decreased by retirements. Therefore, the graduation rate can be seen to be the supply in this illustration with no demand feedbacks present in this illustration.

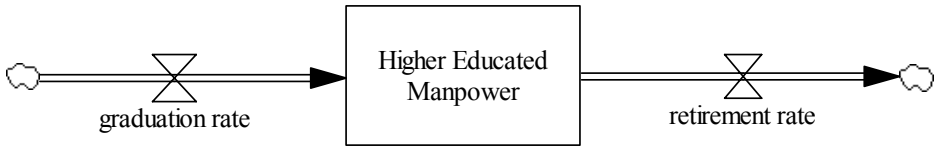


Figure 2: Basic Higher Educated Manpower Illustration

In essence it is this "graduation rate" the government is trying to increase, to 50% of all 18-30 year olds. This means that any model(s) built will have to contain all 18 year olds to "retirement age". Further investigation into the arrant status of 18 year olds shows that the majority are at colleges of further education, which is taken to include 6th forms or 6th form colleges. Further education (FE) is the first post-compulsory educational institute that people can choose to enter into. It is also the phase some adult leavers need to complete before enrolling at university. This means that FE will be a level in the model. This level will need to be fed by a flow rate. With the majority of FE students coming from high schools, which is the last compulsory educational

institute, it would seem appropriate and sensible to have "high school leavers" as an inflow into the model. The boundary has thus been pushed back even further to include 16 year olds.

The government's targets are only for "home students" and as modelling the graduation rate would incorporate both EU and Non-EU students (the definitions of these have been shown below) it would also be sensible to have these groups held separately. On a further note about the classification of the students (and later employees) in the model, it is also sensible to separate out males and females. This is often carried out in governmental reports and one aim of this model and research is to keep as closely as possible to the terminology and characteristics of UK government papers.

Home, EU and Non-EU Student Definitions:

- Home students: Any student from the British Isles. Traditionally, home students have not been profitable for universities as they pay little towards their expense, though this is due to change in September 2006 when universities will be able to obtain almost three times the current fee per student.
- EU students: This is any student from a European Union country. They must pay an international rate for their tuition fees, though funding is sometimes given by their country of origin. They also qualify for "home funding" under certain criteria.
- Non-EU students: These students are predominately from the growing Asian economies and are the most profitable for the university. Their tuition fee criteria are much stricter than for the EU students, but their numbers have continued to show strong growth, and are predicted to remain this way.

The model will therefore contain, and keep separate:

- Males and females
- Home, EU and Non-EU Students

There are two ways of achieving this. The first is by building an individual model for each group and the other is by using subscripts. The model built was built using subscripts (3 x 2 = 6 in total). This option leaves the possibility of adding more subscripts later, which will be especially useful for looking at individual subject areas. An example of this has been illustrated below.

Undergraduates At UK Universities [population] =… where [population] covers all six categories mentioned above (home male, home female, EU male, EU female, non-EU male, non-EU female). This equation accumulates the numbers in UK higher education at any one time, having regard for the flows in (3 inflows) and possible flows out (7 outflows).

To help in deciding where variables fit into the model a Three E's Table (TET) is created (table 1), which is created by taking a boundary chart (figure 3) and converting it into a table format.

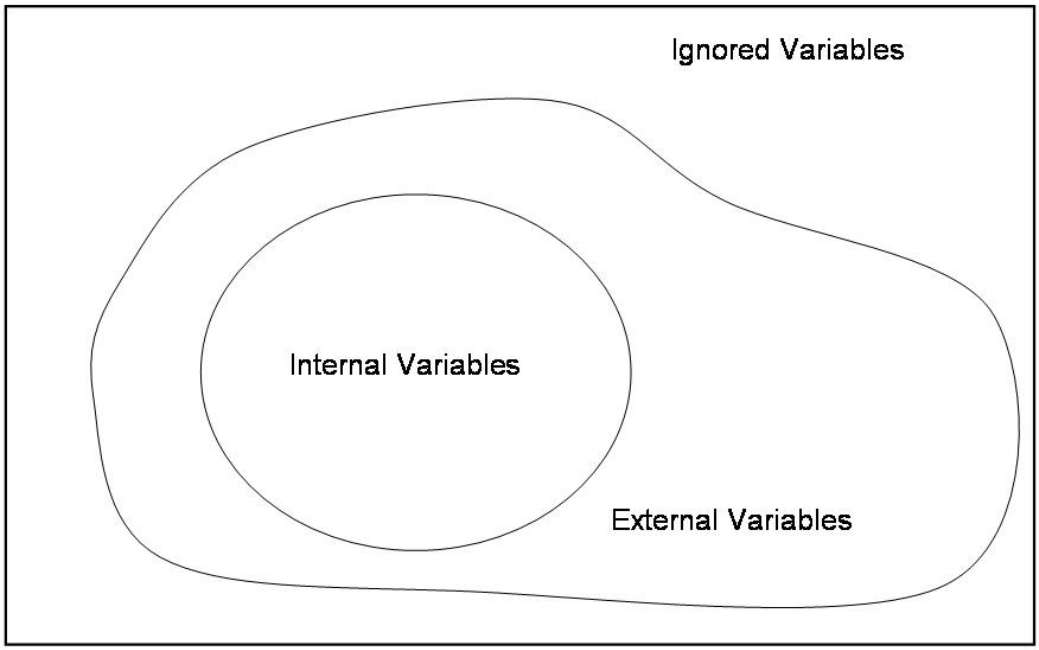


Figure 3: Boundary Chart

A model boundary chart allows variables to be classified, dependent on whether they are internal, external or ignored. In SD terminology internal variables are variables that are endogenous and will be affected by changes in the model, through feedback loops. The external variables are exogenous, as they are built into the SD model, but are not affected by the models behaviour. Ignored variables are classified as external variables, which either have little or no affect on the model or are not included as they are integral to the objectives being studied. All variables are usually explained in more detail.

Endogenous Variables	Exogenous Variables	External Variables
Undergraduate population in the UK	High school leavers	Number of university admin staff
Postgraduate population in the UK	Immigration in to the UK	Grades of graduates
Skilled population in the UK		

Table 1: Example of Three E's Table Used in the Model Building Process

Supply Sector

Supply of skilled workers comes from graduate students and postgraduate students, and the model will need to show where the undergraduates and postgraduates come from.

As mentioned in the previous section, the model has high school leavers as an inflow into the model. This is not a single inflow as the high-level map shows that high-school leavers can go into employment, unemployment or FE. These three inflows require data to be read in (the author has chosen to read this data in from a Microsoft

Excel™ spreadsheet rather than ".dat" file to increase the ease of possible changes, though steady state in the percentage splits, for these three flows, has been evident for a number of years). The model has been designed to run from 1996 to 2020. The 24 years of past data/predictions has been derived from birth records and projected forward to predict the numbers of high school leavers (figure 4). This is possible since high school leavers up to 2020 will have been born up to 2004.

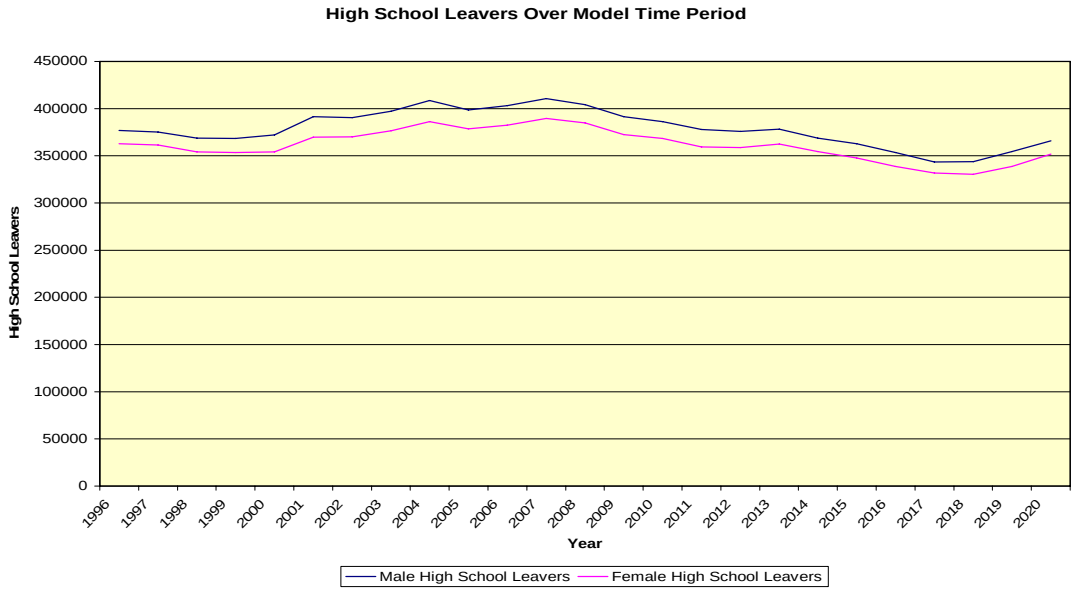


Figure 4: The Expected Number of High School Leavers over the Model's Time Period.

By analysing trends in the choices students make i.e. FE, employment or neither, this data was divided (in the Excel sheet) and imported into the corresponding inflows. The dividing of the data may be integrated into the SD calculations, rather than imported into the model, in future work.

Because this section looks at the supply of higher educated manpower, the high school leavers who have carried on in further education are the only ones that relevant. The FE students spend an average of 2 years in college (unless they dropout) before entering into employment, unemployment or to enrol on an undergraduate course. The percentage enrolling at university stood at 42% in 2003 (Brown 2004) and is targeted to hit 50% by 2010. After graduation from the undergraduate level, the choice is the same for the graduates as it is for the students finishing either high school or further education, with one exception. University graduates can also enter the skilled employment sector. The skilled employment sector comprises of the jobs that require the employee to have at least a bachelor's degree. The size of this sector is often contested by economists and academics with some believing that the market may already be saturated (Pearson 1993) and those that believe 80% of the new jobs created between 2004 and 2010 will be skilled employment (Clarke 2003).

The discussion so far has concentrated on home students. All of the high school leavers and undergraduate enrolments so far discussed have been home students. EU and Non-EU students enter the model at two distinct points: as undergraduates and postgraduates. They study for a shorter length of time than home students, because they do not take "sandwich years" or dropout to change course (due to their visa obligations and cost involved). EU and Non-EU graduates do not enter employment

immediately and instead return home. EU and Non-EU graduates do not directly enter into the employment sector. After graduation these students return home. For EU and Non-EU students to enter employment in the UK they must emigrate. The emigration figures have still to be included in the model and this is discussed more in the next chapter. Figure 3.4 shows the numbers of students who have entered undergraduate degrees from 1996-2004 and the projected figures from 2005-2020, that are used initially in the model. Explorations of variations in these projections are anticipated. Currently the future projections are worked out from a four-point moving average.

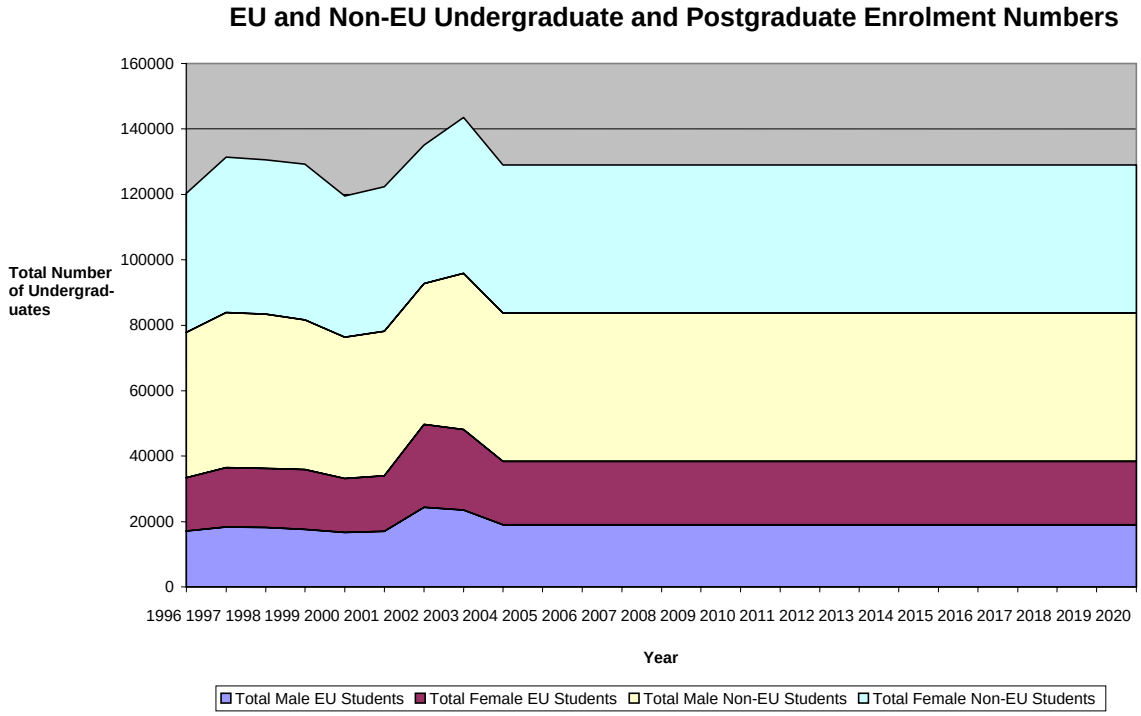


Figure 5: Graph of EU and Non-EU Undergraduate and Postgraduate enrolment numbers.

An increase in both undergraduates and postgraduates, additional to those continuing on from compulsory education, comprises those that have returned to HEI after employment. These are usually adult-learners that carry out courses part-time (though, not exclusively) to gain an advantage in their workplace, better themselves or change career paths. This is show on the model views later but is not active in the model. Therefore, the supply of skilled labour can be seen to come only from undergraduates, postgraduates and migration. Unskilled labour can come from all of these sources with the addition of colleges and high schools.

Demand Sector

In recent years the number of papers and books trying to answer what level of demand there is for skilled employees, in the UK, has risen sharply. The majority of these studies use the USA's rapid growth of skilled labour and under-utilisation as an example of what could happen in Britain. However, all papers agree that HEI are the places where skilled employees should be educated and trained. HEI are increasingly becoming their place of employment for skilled employees (First Destination Statistics, Annual).

The demand for higher educated manpower comes from the "skilled employment" level in the model. This level contains the graduates that have successfully acquired their degrees and entered into employment which makes use of their qualifications.

This is the section of the model that needs to be enhanced and improved over the next six months and is discussed more in the "future work" section. However, the initial literature review indicates that the demand does not affect the subject areas that people study as, in general, there seems to be a large over-education of the employment sector already.

The literature datasets (i.e. education statistics annual reports) show that some subject areas have excessive supply. In general larger subject areas (e.g. business studies, media, computing) are "supply push models" rather than "demand pull models" – which will most likely be in areas such as the sciences, medicine, nursing and dentistry. There are also "caps" and levels which affect the throughput of the model (e.g. the number of places available for students hoping to become doctors).

The Model

There paper does not go into detail about the construction of the model although views of the model structure have been given in figure 6, figure 7 and figure 8.

The first view is of "further education and undergraduates" (figure 6).

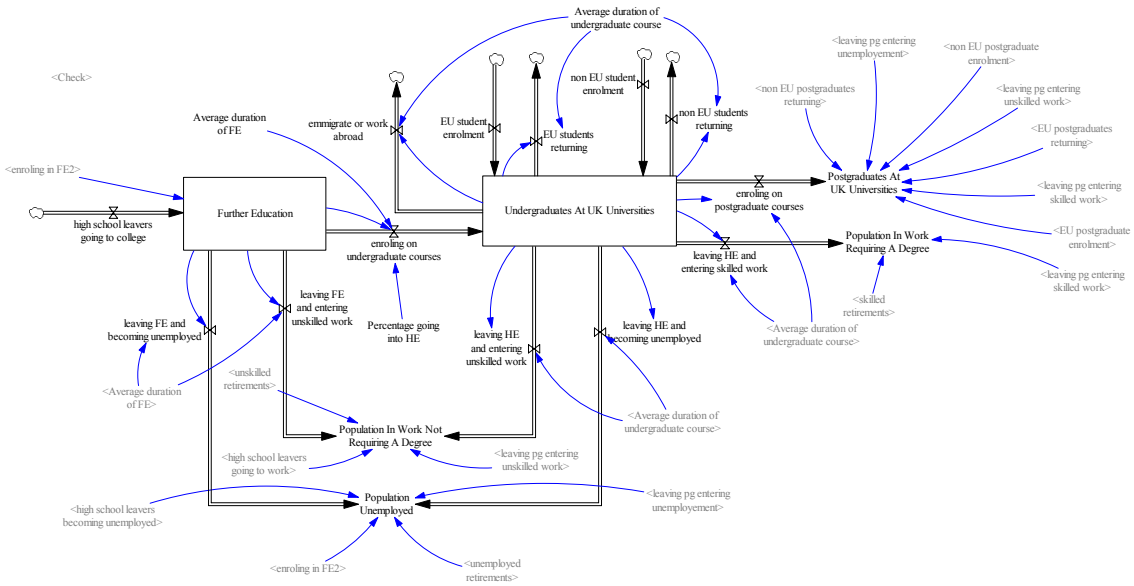


Figure 6: Further Education and Undergraduates View

This view (figure 3.5) shows the flow of high school leavers into further education (FE). They remain in FE for 2 years before entering either employment (unskilled), unemployment or enrolling at university. The current model divides the split as follows (all subjects are "home"):

- Enrolment (males and females 43%)
- Employment (males 39%, females 44%)
- Unemployment (males 18%, females 12%)

These numbers are taken from the annual statistical reports on education and employment. The high number entering unemployment is due to "gap years". This is the highest proportion of the population that is classified as unemployed. The ratio of females gaining employment is above that of males, in part due to females accepting part-time work positions more so than males.

Visible on this are the undergraduates from EU and Non-EU sources, and because the model structure is the same for males and females they use the same "pipeline flows" and are affected by the same information.

Emerging from the "Undergraduates At UK Universities" is a number of flows. If the EU and non-EU flows are ignored these split into two main categories: flows entering employment and the flows entering education. The employment flows are:

- leaving HE and entering skilled work;
- leaving HE and entering unskilled work and;
- leaving HE and entering unemployment.

And the education flows are:

- emigrating or work abroad and;
- enrolling on postgraduate courses

Figure 7 compares the model output to actual data for the Further Education students. Note that actual data only runs from 1997-2003 (1996 figures are still to be included and compared against).

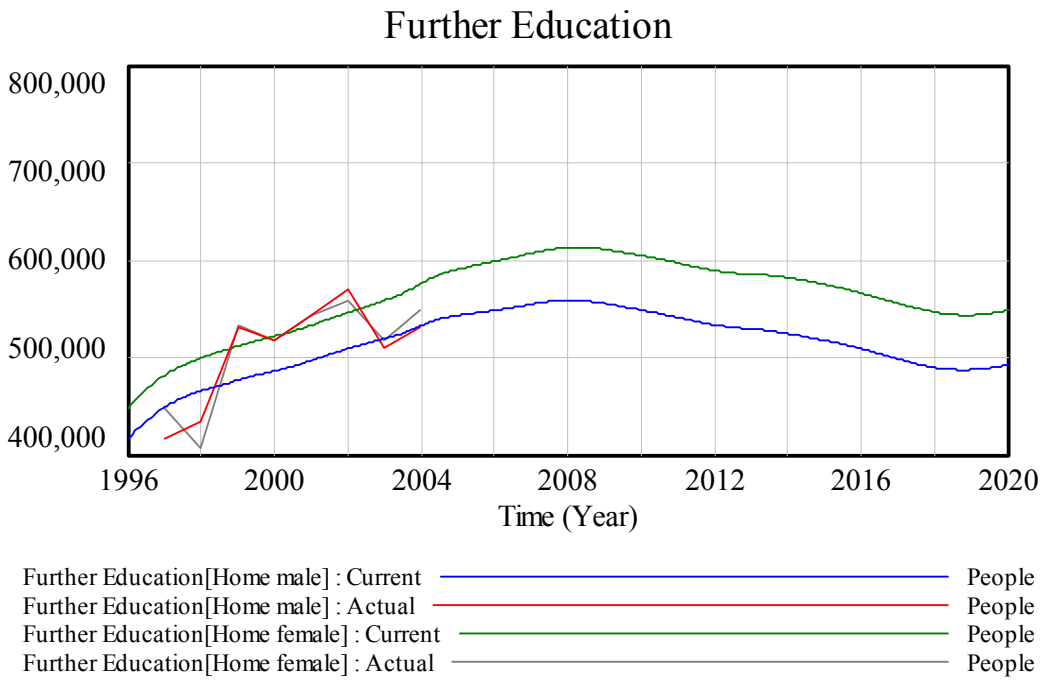


Figure 7: Actual Data Compared to Model Data for FE Students

The details for postgraduate study and employment have been placed on separate views so not to clutter the view. The next view examined is the "Postgraduate" view (figure 8).

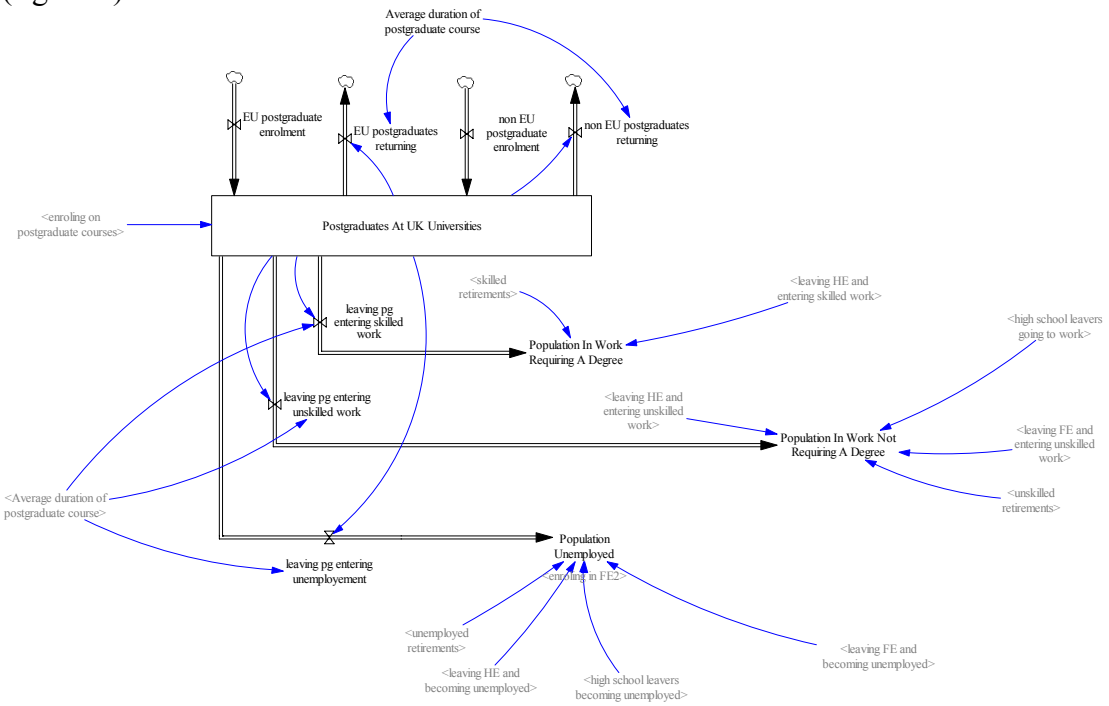


Figure 8: Postgraduates View

The "Postgraduates" view (figure 8) follows a similar model structure to the undergraduates. There are different changes in the length of study and percentages going into skilled jobs, but the options remain the same. There are a higher proportion of postgraduates entering skilled employment, which may be due in some part to a higher proportion entering academia, though further work is required over the second half of the research to investigate this. There is also no further study (expect job training) higher than postgraduate level so the options are also limited, again adding to the increased percentages.

Both the undergraduate and postgraduate views are linked to the employment view. This section is underdeveloped and will be one of the larger extensions in the next year. The "Employment" view (figure 9) displays all of the outflows from the undergraduate and postgraduate views, as inflows, alongside the additional inflows of "high school leavers entering unskilled work" and "high school leavers entering unemployment".

The only exit flows from these levels are due to retirements, at which point people leave the model entirely. The other outflow is a single outflow that is currently non-operational, and will be used to model part-time FE students as this view progresses.

There is also an additional views (omitted from this paper) that check and validate the model.

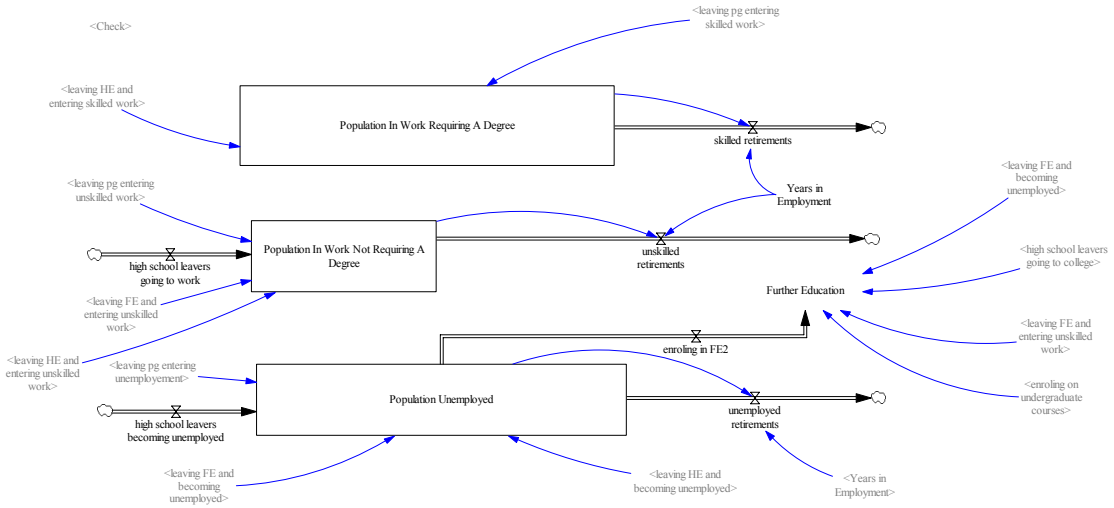


Figure 9: Employment View

Future Work

The model that has been built needs improving, whilst replicating aspects of this original model to build subject specific model(s).

The "University Model" requires two key areas of improvement. The first is to consider limits on the stocks and flows. The current model allows for unlimited resource flows, and is not representative of the real life situation. For example, there are only a certain number of skilled jobs available. Though the model limits the percentage of employees per time period entering the model, it does not limit the number. By placing limits on these flows it will highlight the dynamic behaviour of the flow of possible employees, as they fail to find skilled work and enter unskilled work. Many case studies and papers, mentioned in the literature review, highlight this factor and it is important to make this explicit in the framework of the model.

Also, the model needs to include part-time students. This is an important short-term aim. Such an improvement would be possible in a relatively short time span and add extra dimensions and flows to the model. However, care must be taken not to overcomplicate the model. In addition, improvements have been made to check the model's output and improvements have been made the author's modelling technique.

The model is currently pitched at the national level, modelling all students and all courses under broad terms. By splitting students into smaller sub-sets or devising new models into which individual course and subject areas can be placed makes for a more detailed analysis of trends and possible future education levels in employment can be obtained.

There are certain areas that have already been viewed as particularly important. These areas are medicine, nursing and dentistry. By building a model to look at these subjects any skills shortages will be highlighted in a manner not possible with the broader perspective model. Each subject area will have its own individual dynamic behaviour and characteristics, so there are only a limited number of models that it will be able to construct in the time available frame. Diversity in the models, but concentrating on key public issues, will be the guiding aim as these models are created.

Glossary

College – refers to any institute of further education.

Demand Pull Models – models that depend on the demand pulling through entities from the possible supply roots in order to satisfy demand (often demand outweighs supply).

Home Students – residents of the United Kingdom (UK) or students holding a UK passport.

Knowledge Based Economy – is an economy that gains economic benefit through high-technology businesses and intangible products, such as innovative ideas, and also through educational and research institutions' output. Sometimes knowledge based economy is referred to as "**knowledge economy**" or "**knowledge driven economy**".

Production Based Economy – is an economy that is largely based around the production of tangible goods. Often, production economies are reliant on low wage workers and the country's natural resources, e.g. in mining or logging. Sometimes production based economy is referred to as "**production economy**" or "**production driven economy**".

Retirement Age – this is 65 for males and 60 for females, in line with current government retirement ages. Any exemptions have been artificially set as these ages too.

Sandwich Years – an optional year, usually toward the end of an undergraduate course, in which a student can take a year's placement in suitable employment.

Skilled Employment – requiring people with bachelor's degrees or higher.

Supply Push Models – models where the supply chain pushes through output to try to satisfy demand.

University – In this text, the terms "university" or "**universities**" is used to cover all higher education institutions, such as universities; university colleges; colleges of higher education; colleges of further education and; other institutions. This is to remain consistent with the terms used by the 'Department for Education and Skills' and in particular their 2003 White Paper on 'The Future of Higher Education'.

Unskilled – not requiring a bachelor's degree.

References

Alpin, C., Shackleton, J. R. and Walsh, S., (1998) Over- and Under-education in the UK Graduate Labour Market, *Studies in Higher Education,* vol. 23.

Batty, P., (1999) A Piper Who Does Not Play The Player's Tune, *Times Higher Education Supplement.*

Brown, P. and Hesketh, A., (2004) *The Mismanagement of Talent: Employability and Jobs in the Knowledge Economy*. Oxford: Oxford University Press.
Clarke, C., (2003) The Future of Higher Education. IN DFES (Ed.), Crown Copyright.
Dangerfield, B., (2005) Towards A Transition To A Knowledge Economy: How System Dynamics Is Helping Sarawak Plan Its Economic and Social Evolution IN Sterman, J. D., Repenning, N. P., Langer, R. S., Rowe, J. I. and Yanni, J. M., (Eds.) (2005) *23rd International Conference of the System Dynamics Society.* Boston, USA, The System Dynamics Society.
Kennedy, M. C., (1999) Some Issues in Building System Dynamics Model for Improving the Resource Management Process in Higher Education, *17th International System Dynamics Conference.* Wellington, New Zealand, System Dynamics Society.
Mason, G., (1996) Graduate Utilisation in British Industry: the Initial Impact of Mass Higher Education, *National Institute Economics Review,* pp. 93-101.
Pearson, R., Seccombe, I., Pike, G. and Connor, H., (1993) Employer Demand for Doctoral Social Scientists?, *Studies in Higher Education,* vol. 18.
Robinson, P., (1995) Qualifications and the Labour Market: Do The National Education and Training Targets Make Sense?, Centre of Economic Performance (Working Paper).
Statistics, F. D., (Annual) Universities Statistical Record and Committee of Directors of Polytechnichics. IN CNAA (Ed.) *Annual Reports.*
Thurow, L. R., (1972) The Economics of Overeducation, *Economics of Education Review,* vol. 4**,** pp. 93-104.
Wilson, R. A. and Green, A. E., (2001) Projections of Occupations and Qualifications: 2000/2001 Research in Support of the National Skills Taskforce. Sheffield, Department for Education and Employment.

LVT Position Sensor for Multifinger Teleoperation System

Amran Mohd Zaid

School of Computing, Science and Engineering

Introduction

Recently there are several types of sensors using in multifinger robot hand such as bend sensor, encoder, potentiometer and Hall Effect Magnetic Sensors (Hall Effect Magnetic Sensors Based on Polysilicon TFTs) (Carvou et al. 2004)

Bend/flex resistor

When the bend/flex sensor (figure 1) is bent, its resistance will increases. The increase in resistance can be used to determine the finger position. Unfortunately, the bend/flex sensor suffers from high hysteresis.

Figure 1: Bend sensor

Encoder

Two types of encoders are commonly available; **Incremental Encoders** and **Absolute Encoder.** Incremental encoders are ideal for angular or linear position sensing. When integrated into a control system, incremental encoders provide motion and positioning feedback data to the systems controller. Incremental encoders are also effective when used to digitally measure length or position by counting pulses related to the unit of length of travel. Absolute encoder can be divided into two types i.e. **Singleturn Absolute Encoders** and **Multiturn Absolute Encoder.** Absolute encoders provide a unique bit output for each resolvable position. Absolute encoders are typically used in applications requiring precise axial positioning. Available output codes include Gray, Gray-Excess, Binary and BCD, while different serial code outputs permit interfacing with PLCs (Peperl andFuchs Worldwide 2006).

In a multifinger teleoperation system, an encoder's tendons routing has to go through translational and rotational motion, which may cause delays when the finger returns to the rest position. Encoders also occupy more space which made the system look bulky.

Potentiometer

Potentiometer expose to wear and tear relatively fast. Two major type of potentiometers are available i.e. slide potentiometer and rotary potentiometer.

There are several types of rotary potentiometer such as moulded case rotary potentiometer, metal shaft rotary potentiometer, single unit rotary potentiometer, dual units rotary potentiometer, multiple units rotary potentiometer, insulated type rotary potentiometer, rotary remote control potentiometer, snap-in rotary potentiometer etc (Electric 2006). In robotic finger application, very frequently rotary potentiometer was used as a position sensor.

Normally the wiper of a potentiometer will be attached to the finger joint. When the tendon pulls the finger, it will move the wiper. By applying a constant voltage to the end terminals of the potentiometer, the wiper voltage will change as a result of the finger movement. Although simple, simple to use, and inexpensive, most of the problems arise from the mechanical contact between the wiper and the resistive layer and the mechanical linkage between the finger and the wiper. In order for the contact to withstand even moderate levels of vibration, a spring pushes the wiper against resistive layer. The force of the spring is determined by two very contradictory requirements (Landmann):
a. This force has to be as strong as possible to maintain the contact under vibration or high acceleration,
b. This force has to be as small as possible in order not to wear the resistive layer. Clearly any compromise will result in a device not able to withstand harsh environments or with a poor reliability.

Hall Effect Sensor
Magnetic Position Sensors include digital and analogue Hall-effect position sensors, Hall-effect vane sensors, Hall-effect basic switch, and magnets (Honeywell 2006). These sensors respond to the presence or the interruption of a magnetic field by producing either a digital or an analogue output proportional to the magnetic field strength. Digital and analogue "sensor-only" devices are operated by the magnetic field from a permanent magnet or electromagnet. The actuation mode depends on the type of magnets used. Integral magnet position sensors are operated by either a vane passing through a gap or a magnet mounted on a plastic plunger.
Early Hall-effect sensor designs utilised a single Hall element (figure 2), while many designs originated within the last ten years utilised a four-plate Hall-element array (figure 3), which can be considered a resistor array similar to a Wheatstone bridge (Gilbert 2006). The quadratic array placed four Hall-plates in parallel providing a “mechanically averaged” Hall voltage. Offset errors and mechanical stresses tended to cancel out; nearly a 10x improvement in stability and stress immunity can be realised using this scheme (figure 4).

The problem with the Hall Effect sensor if we were to use this as a position sensor in the multifinger dexterous manipulator is the non linearity. Highly non linearity effects in Hall Effect sensor will create difficulty in the control part.

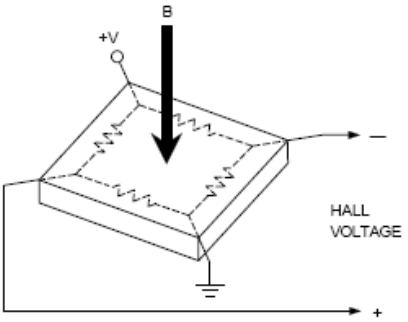


Figure: 2: Single Hall-plate

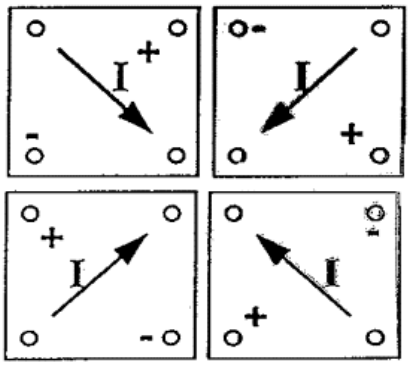


Figure 3: Quadratic Hall-plate

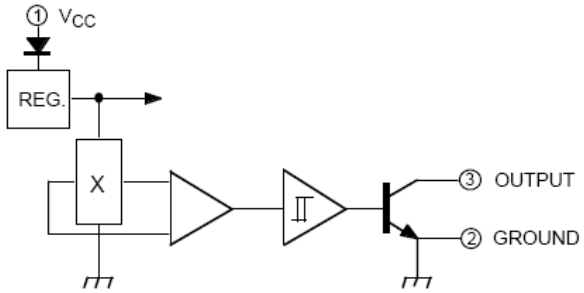


Figure 4: Second generation digital sensor

Magnetic Theory

The basis of Faraday's law

Faraday's Law (R.Clarke) states that a voltage, e, is generated by a coil of wire when the magnetic flux, Φ, enclosed by it changes and is given by:

$$e = N \times d\Phi/dt \quad \text{volts} \qquad \textbf{Faraday Law} \qquad \textbf{------ (1)}$$

Where N = number of turns
Φ = magnetic flux in Weber
t = time

It doesn't matter what causes the flux change; the result is an induced voltage, and the faster the flux changes the greater the voltage.

Generally if Φ = instantaneous value of flux in Weber
= $\Phi_{pk}\ Sin2\pi ft$

therefore instantaneous value of induced emf per turn in dΦ/dt volts (Hughes, 2002)

$$\frac{d\phi}{dt} = 2\pi f\phi_{pk} Cos 2\pi ft \qquad \text{Volts}$$

$$= 2\pi f\phi_{pk} Sin(2\pi ft + \pi/2)$$

therefore maximum value of induced emf per turn = $2\pi f\Phi_{pk}$ volts
and r.m.s. value of induced emf per turn is:

$0.707 \times 2\pi f\Phi_{pk} = 4.44f\ \Phi_{pk}$ Volts

Hence r.m.s. value of primary emf is

$e_{1rms} = 4.44N_1 f\ \Phi_{pk}$ volts

and r.m.s. value of secondary emf is

$$e_{2rms} = 4.44N_2 f \Phi_{pk} \text{ volts}$$

We can easily generalise this for any flux varying sinusoidally at a frequency f to show:

$E_{rms} = \sqrt{2}\pi N \times f \times \Phi_{pk} = 4.44\ N \times f \times \Phi_{pk}$ volts **The transformer equation**

$$\Phi_{pk} = \frac{e_{rms}}{4.44\,fN} \qquad \text{-------} \qquad (2)$$

By integrate equation (1) we get the integral form of Faraday's Law,

$\Phi = (\int_0^t e.dt) / N$ webers -------(3)

where e is the externally applied voltage and N is the number of turns.

Inductor with AC applied

Let's apply a sinusoidal voltage, frequency f, RMS amplitude a -

$e = (\sqrt{2})a.\sin(2\pi.f.t)$

Substituting this into equation (3), the integral form of Faraday:

$\Phi = (\int_0^t (\sqrt{2})a.\sin(2\pi.f.t)\ .dt) / N$

$\Phi = ((\sqrt{2})a/N) \int_0^t \sin(2\pi.f.t)\ .dt$

$\Phi = (-(\sqrt{2})a/(2\pi.f.N))\ [\cos(2.\pi.f.t)]_0^t$

The expression with the limits of integration will always be between -1 and +1 so that the peak value of flux is given by

$\Phi_{pk} = (\sqrt{2})a/(2\pi.f.N)$

$\Phi_{pk} = a / (4.44\ f.N)$ Wb

Compare this with the equation (2) above.

LVT Sensor

An LVT sensor is self-contained and requires minimum mechanical components. This eliminates the noise normally associated with mechanical swiping and rubbing allowing for greater accuracy than with traditional rotary and linear potentiometer. Furthermore, the LVT sensor is not prone to degradation through mechanical contact, so it has a longer life than many competing sensors. The longer life of components means fewer system breakdowns due to mechanical failure, so it is more reliable and less expensive to the system operations.

Construction of LVT Position Sensor

Four major components are involved in the construction of the LVT sensor i.e.:

coated enamel wire 0.15mm diameter

rolled paper

extension spring (4.5mm external diameter)

adhesive

The primary and secondary coils of the sensor were wound on to a 5 mm diameter rolled paper. The coils are from the coated enamel wire of 0.15mm diameter and have 40 turns each. After completing the winding process, the coil is fixed and hardened with "impact adhesive" and covered with another layer of paper. Figure 5 illustrate the construction of the sensor. The rolled paper complete with primary and secondary

winding (and the finger spring inserted into it) is fixed to the slot wall of the finger as shown in figure 6. Inserting the LVT sensor into the slot wall of the finger didn't need extra space for the system, hence will make the system more compact.

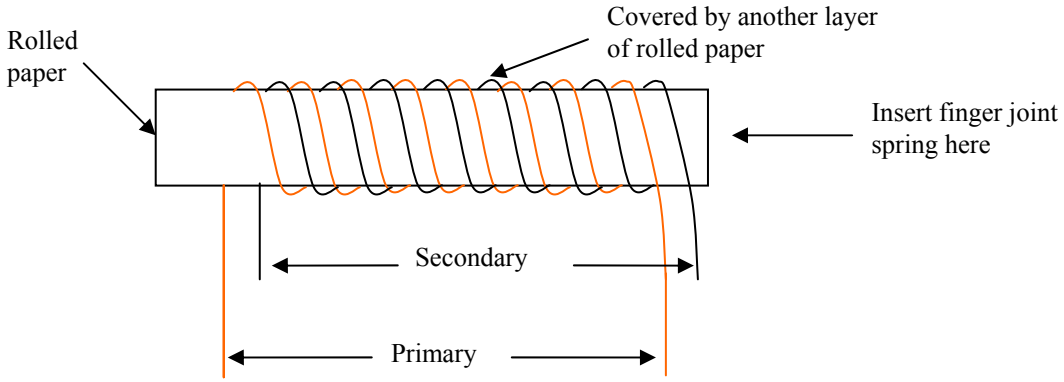


Figure 5: LVT Position sensor

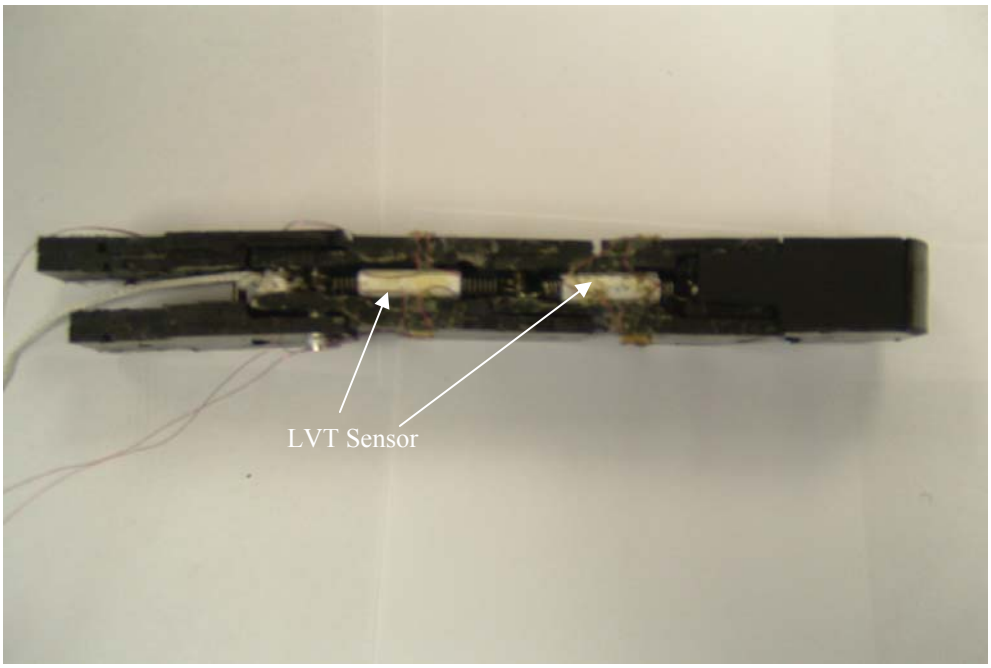


Figure 6: LVT sensor fixed to the slot wall of the finger

LVT Sensor Operation

When a sinewave signal from a signal generator is supplied to the primary winding of the sensor, a magnetic field is formed surrounding the coil and inducing a sinewave signal to the secondary coil.

The finger spring inserted into the LVT sensor acts like a core. When the spring is extended as the tendon pulls on the finger, the "core" become less solid, the magnetic field will become weaker hence decreasing the induced voltage in the secondary coil. On the other hand if the spring is compressed as the tendon releases the finger joint, the "core" becomes solid, and the magnetic field will become stronger, hence increasing the induced voltage in the secondary coil. With the calibration process, the

variation of secondary voltage (as tendon flexs or extends the finger) can be used to determine the finger position.

Amplification of LVT Sensor Output Signal

In this project experiment, we found that using 50 kHz, 125 mV_{p-p} sinewave signal supplied to the primary coil will be the optimum value to get undistorted output for our application. The sensor output signal in the secondary winding is 50 mV_{p-p} when the finger in rest position. To make the output a useful signal, we need to amplify this signal. The amplification circuit board has been developed to fulfil the need. The board consists of an Instrumentation amplifier AD620AR, rectifier and Rail to Rail High Output Current Quad Operational Amplifier TS924. Sensor output signal is fed to the AD620AR giving an amplified AC output signal. This amplified output from the AD620AR is converted to a DC signal at the rectifier stage and the DC will proceed to TS924 amplifier to get a final amplified voltage (as illustrated in figure 7). With the amplification circuit board, we get an amplified sensor output that varies in the range of between 1V to 2V. This forms an acceptable feed signal to the PID controller board and provides good resolution of control.

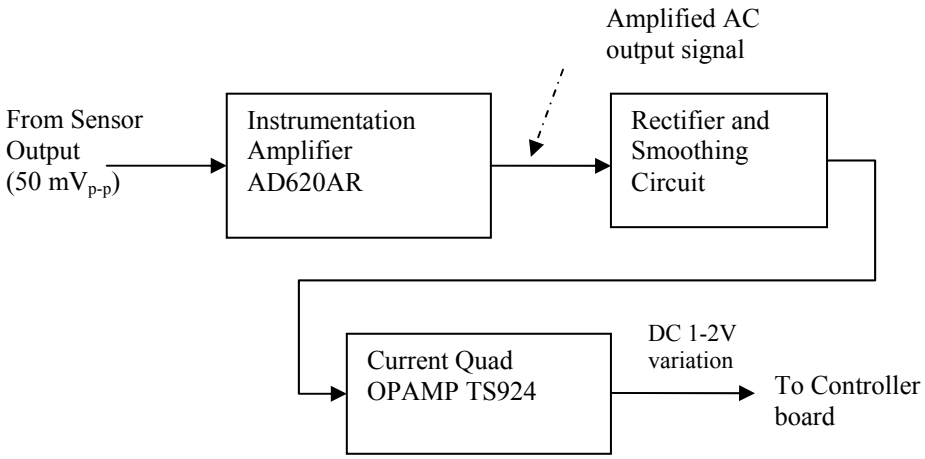


Figure 7: Amplification board block diagram

Figure 8 shows the schematic diagram of the sensor output amplifier circuit.

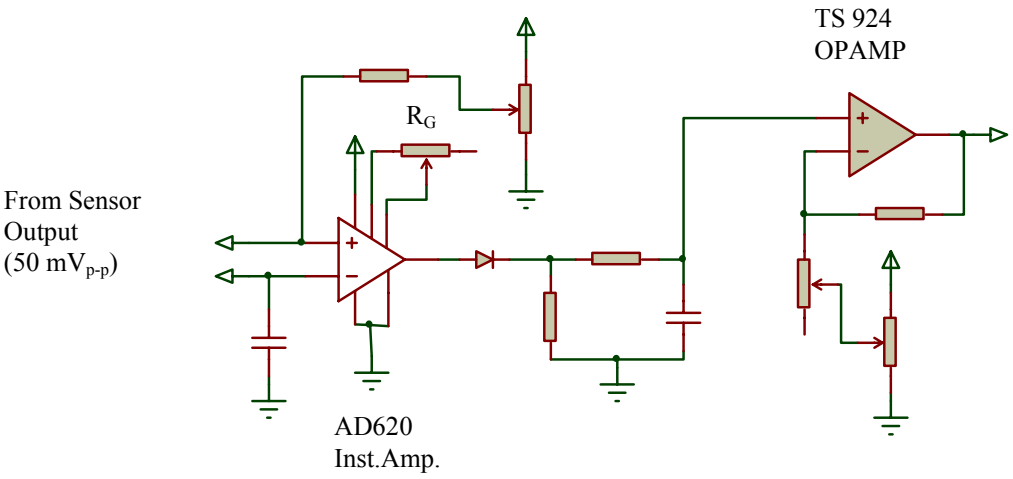


Figure 8: LVT sensor output amplifier circuit

Results

Maximum Magnetic Flux

In this project, an LVT primary coil have 40 turns, a sinewave signal frequency of 50 kHz and a peak to peak primary voltage Ep-p of 125 mV were used.

From equation (2), looking at the primary coil, the peak value of magnetic flux is given by:

$$\Phi_{pk} = \frac{e_{rms}}{4.44\,fN}$$

where N = 40 turn
f = 50 kHz
Ep-p = 125 mV

$$e_{rms} = \frac{E_p}{\sqrt{2}} = \frac{E_{p-p}}{2\sqrt{2}} = \frac{125x10^{-3}}{2\sqrt{2}} = 44\,\text{mV}$$

$$\Phi_{pk} = \frac{44x10^{-3}}{4.44x150x10^{3}\,x40} = 1.65\ \text{nWb}$$

Amplification Board

With the input 50 mVp-p, resistor R_G (resistance between pin 1 and 8 AD620AR) has to be adjusted to give an undistorted output. From the experiment, when R_G is set to 479 Ω, AD620AR is given maximum 5.2 Vp-p output without distortion. Therefore maximum gain without distortion is:

G = 5.2V/50mV = 104

Alternatively, from the data sheet, gain for AD620AR is given by:

G = 1 + [(49.4k)/R_G] ; Where R_G = Resistance between pin 1 and 8

$= 1 + [(49.4k)/479]$
$= 104$

To have an instrumentation amplifier output without distortion, the instrumentation amplifier gain AD620AR must be less than 104. In order to achieve this, R_G must be adjust to the value of greater than 479 Ω.

In this project, when the 5.2Vp-p AD620AR output (finger at rest position) passes through rectifier circuit, AC voltage is converts to around 1Volt DC voltage when finger in rest position. The DC voltage will change to 0.7 volt when finger was bent to 90^o. In other words, the rectifier DC output voltage will fluctuate between 0.7V to 1V as the tendon flexs and extends the finger joint. This fluctuating DC voltage is amplified by the TS924 OPAMP, giving a final variable DC voltage to be fed to PID controller board to perform control action. Figure 9 showing that the final amplification circuit board output voltage is 2.4V when the finger in rest position and reduces to 0.8V when the finger is bent to 90^o (figure 10). We can clearly see that the output of amplifier fluctuates between 2.4V (finger in 0^o position) and 0.8V (finger in 90^o position) giving 1.6V amplitude of variation.

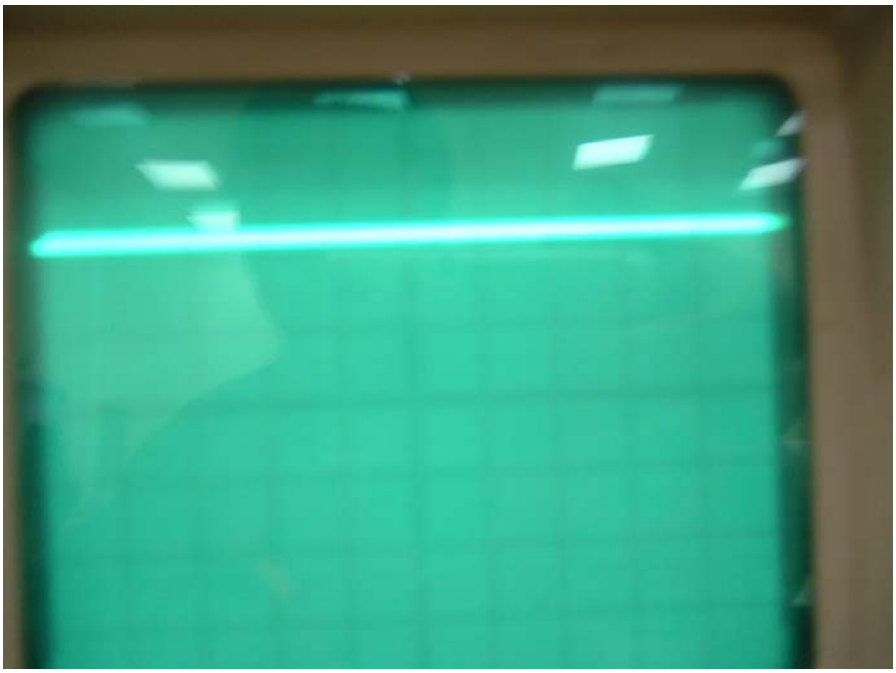


Figure 9: Amplification circuit board output voltage when finger in rest position, DC output 2.4V

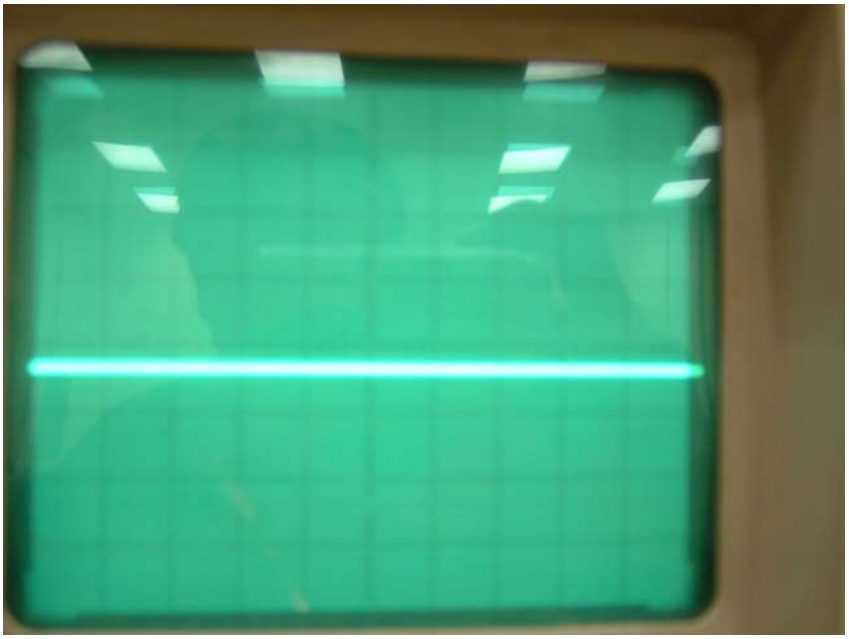


Figure 10: Amplification circuit board output voltage when finger was pulled to 90^{o} position, DC output reduced to 0.8V

Hysterisis Test for LVT Sensor

A hysterisis test was run to observe the repeatability of the LVT sensor. The LVT sensor output was compared to a potentiometer output and the data was recorded. Figure 11 shows the block diagram on how the hysterisis test was conducted. Both the LVT sensor and the potentiometer were coupled to the finger joint. The potentiometer was fed by 5V DC supply while the LVT sensor was fed by a sinewave signal. When the tendon pulled the finger joint, potentiometer and LVT will move simultaneously and produce outputs. Outputs from both the potentiometer and LVT were fed to the ADC multiplexer in the microcontroller ATMEGA 128 ("ADC input 0" for potentiometer and "ADC input 1" for LVT). By using a serial RS232 link, the data collected from the microcontroller was send to the computer, and Matlab Simulink was used for the data acquisition process.

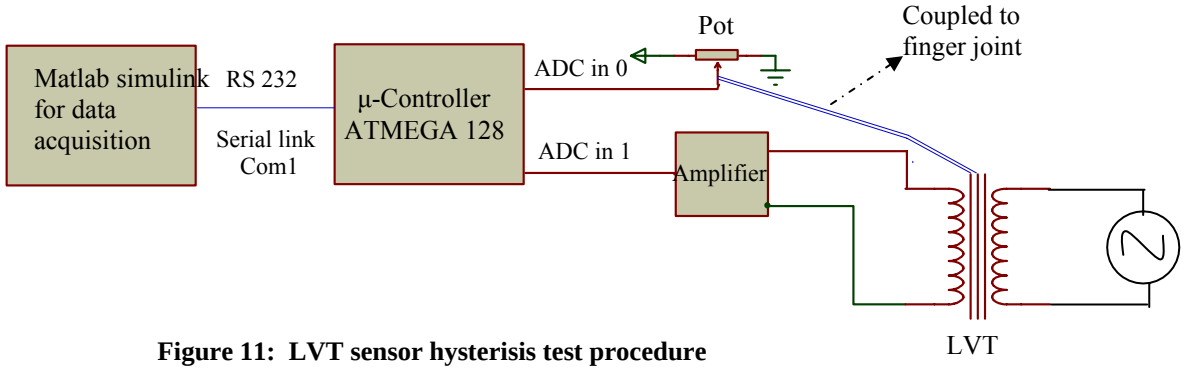


Figure 11: LVT sensor hysterisis test procedure

While the tendon was flexing the finger joint, the output voltages from the LVT and the potentiometer were plotted by the Matlab Simulink. We found that the LVT output follow the pattern of potentiometer output as shown in figure 12.

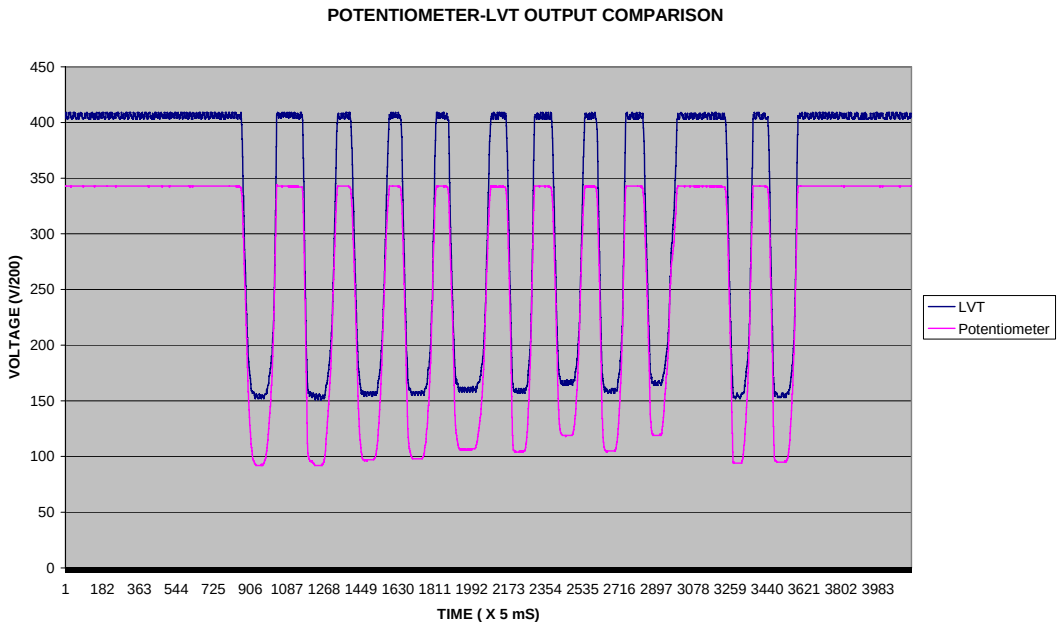


Figure *12: Potentiometer- LVT output comparison*

Hysterisis plots for the LVT sensor also show that the sensor has low hysterisis loss and good repeatability. Figure 13 shows the plot of a low LVT hysterisis loss. From the plot, we can calculate the maximum hysterisis loss as below:

$$\text{Hysterisis loss} = \left(\frac{265-255}{265}\right) * (100\%) = 3.8\%$$

which is less than 4% hysterisis loss.

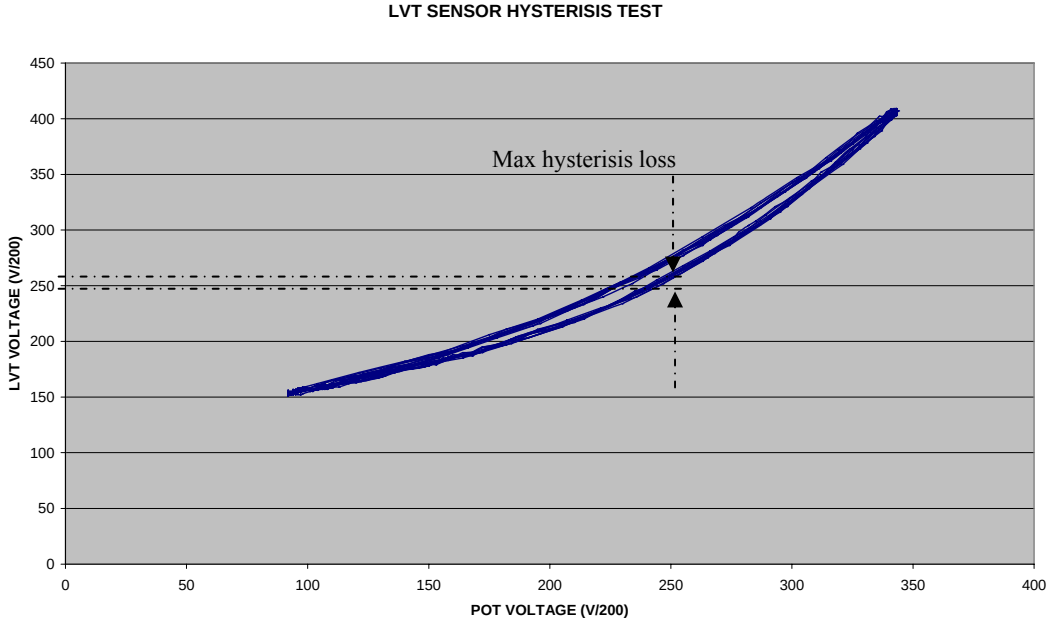


Figure 13: LVT sensor hysterisis test

Conclusion

LVT position sensor is located on the slot wall of the dexterous finger. It could be introduced into this finger joint spring space without requiring any additional space which is an extremely important aspects of sensor design with multijointed robot fingers. The LVT sensor has the added benefits of being self-contained requiring minimum mechanical components. This eliminates the noise normally associated with mechanical swiping and rubbing allowing for greater accuracy than with traditional rotary and linear potentiometer. Furthermore, the LVT sensor is not prone to degradation through mechanical contact, so it has a longer life than many competing sensors. The longer life of components means fewer system breakdowns due to mechanical failure, so it is more reliable and less expensive to system operations. With high repeatability, low hysterisis loss, simple construction, less space and low cost, this sensor promises a better future to replace other sensors which currently are being used in multifinger teleoperation system.

References

Carvou, E., Le Bihan, F., Rogel, R. and Bonnuard, O., (2004) Hall Effect Magnetic Sensors Based on Polysilicon TFTs, *IEEE Sensors 4*, pp. 597 - 602.

Electric, S. H., (2006) Rotary Potentiometer. Song Huei Electric Co., Ltd.

Gilbert, J., (2006) *Technical Advances in Hall-Effect Sensing*. Allegro MicroSystem Inc.

Honeywell (2006) Honeywell Sensing and Control. Honeywell.com.

Hughes, E., (2002) *Electrical and Electronic Technology*. London: Prentice Hall.

Landmann, W., Solid State Replacements for Electro-Mechanical PotentiometerType Pressure Transducers. Advanced Development, Kulite Semiconductor Products, Inc.

Peperl and Fuchs Wolrdwide (2006) Rotary Encoders. Pepperl and Fuchs Designs.

Clarke, R., Induced Voltage: Faraday's Law. http://www.ee.surrey.ac.uk/ Workshop/advice/coils/faraday.html#FYT.