NATURALNESS FRAMEWORK FOR DRIVER-CAR INTERACTION

A thesis submitted for the degree of Doctor of Philosophy

by

Simon Alexander Ramm

Department of Design, CEDPS, Brunel University

ABSTRACT

Automobile dashboards are evolving into intelligent largely screen-based computer interfaces. Recent evidence suggests *unnatural* aspects of some secondary systems (including infotainment systems). *Naturalness of interaction* is a minority discipline not yet applied to the automobile; while *automotive interface research* is a mainly quantitative discipline that has not yet applied a *naturalness approach*. To advance the field, a measurement scale based on rigorous *qualitative exploration of natural-feeling interaction with secondary controls* was required.

Study 1 used *ethnographic interview* with *Contextual Inquiry* inside 12 ordinary drivers' cars, to investigate natural-feeling aspects of past, present and future driver-car interactions. *Thematic analysis* suggested a framework of ten characteristics. Half concerned *control and physicality*; half concerned perceived *socio-intelligent* behaviours of the car.

Study 2 involved intensive *exploratory workshops* with ten drivers comprising *Think Aloud*, *artefact modelling* and *focus groups*, to explore natural-feeling interaction with secondary controls in different ways. The resulting thematic framework comprised 11 characteristics in four categories: familiarity/control, physical connection, low visual/cognitive demand, and humanlike intelligence and communication.

Study 3 comprised two ethnographic *participant observations*. Eight drivers were observed interacting with their controls during long road journeys. Twenty-two drivers were observed interacting verbally with futuristic 'intelligent' secondary systems while driving on public roads. Design guidelines relating to physicality, usability, automation, and humanlike communication were formulated.

Study 4 converted all the qualitative findings into a *questionnaire* comprising 46 bipolar five-point scales. Eighty-one drivers used it to rate one control in their cars. Correlation and factor analyses revealed three underlying factors and 14 items suitable for the *first industrially applicable measurement scale for driver-car naturalness.* These items concern perceived helpfulness, politeness, competence, predictability, control, ease, mental demands, intuitiveness, 'realness', instantaneousness, communication, logical location, mapping and 'affordance'. Initial testing found acceptable validity. The conclusion recommends further data collection, expanded validity testing, and potential applications to *self-driving cars*.

First Supervisor: Prof Joseph Giacomin, Human Centred Design Institute

Second Supervisor: Dr Alessio Malizia, Department of Computer Science

Brunel University, Kingston Lane, Uxbridge UB8 3PH

LIST OF CONTENTS

NATURALNESS FRAMEWORK FOR DRIVER-CAR INTERACTION

CHAPTER 1: INTRODUCTION

1.1 WHY STUDY INTERACTION BETWEEN DRIVERS AND SECONDARY CONTROLS?

From the earliest motor cars to current day luxury saloons with semi-autonomous abilities, it has been essential for their drivers to be able to quickly and easily operate their controls and know the current state of their vehicle through its instrumentation (Harvey and Stanton, 2013). The multiple disciplines comprising the field of *ergonomics* have played a major role in the development of the driver-car interface for over 100 years by optimizing its physical accessibility, comfort, usability and - more recently - the *cognitive ergonomics* of its now sophisticated controls and instruments (Bhise, 2011).

Transport research tends to suggest the car fulfils three basic roles – *instrumental* (i.e. functional), *symbolic* and *affective* (Steg, 2005). However, the latter meaning-based and emotional aspects of driver-car interaction may not always be considered as fully as the functional aspects (Giacomin, 2012a). Cognitive and quantitative approaches to driver-automobile interaction dominate the literature but have been criticised for underestimating human emotions, needs, values and context (Gomez et al, 2004). There are few published studies addressing such aspects of interacting with a car's interface (Angulo, 2007; (Mesken et al, 2007). Many academics also consider the car to have been rather neglected in sociological and anthropological research (e.g. Miller, 2001; Dant and Martin, 2001).

The *primary driving task* (known as 'driving') is generally considered to be a complex multitask activity consisting of interactions between the driver, the car, and the environment, via the *primary driving controls,* requiring the successful integration and coordination of the driver's physical, sensory, psychomotor and cognitive skills (Harvey and Stanton, 2013; Kern and Schmidt, 2009). *Secondary driving tasks* are all the other tasks performed by the driver that are not directly related to the safe motion, speed control and hazard avoidance that characterise the primary driving task. *Secondary systems,* controlled by *secondary controls*, meet the driver's needs for information, signalling, comfort, entertainment and communication (based on Harvey and Stanton, 2013; see Figures 1.1 and 1.2 for examples). Some researchers refer to entertainment and information controls separately as either 'infotainment' or 'tertiary' controls, but in this research they are taken to be a subset of secondary controls.

FIGURE 1.1: COMMON SECONDARY SYSTEM CONTROLS (INPUTS) (FROM KERN & SCHMIDT, 2009)

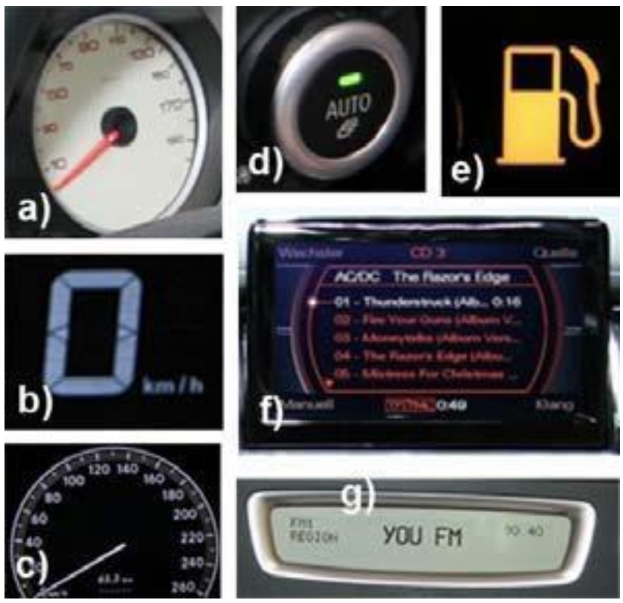


FIGURE 1.2: COMMON SECONDARY INSTRUMENTS (OUTPUTS) (FROM KERN & SCHMIDT, 2009)

Driver-car interaction design has some distinctive and inherent *ergonomic challenges* compared to other human-machine operations. These include designing for users who are usually focused on essential safety-critical ‘primary’ driving tasks rather than ‘secondary’ interactions; the wide variety of potential operating environments in terms of temperature, ambient light and glare; the wide range of possible users and their diverse physical, perceptual and intellectual abilities; the lack of formal training drivers receive in their cars; and the many different use cases cars have to cope with – from long term ownership of one car by one driver, to urban ‘streetcars’ rented by the hour to unfamiliar drivers (Harvey and Stanton, 2013). There are also numerous historical and regulatory design conventions for primary and secondary controls which have to be complied with. Finally, unlike some human-machine interactions, there is a historical expectation that the driver should also *enjoy* interacting with their car and find it satisfying (Harvey and Stanton, 2013).

FIGURE 1.3 EXAMPLE CAR DASHBOARDS FROM 1930 TO 2009 (BMW CARS; FROM KERN, 2012)

Figure 1.3 shows how the driver-car interface has evolved over the last century. The earliest car instruments of 100 years ago were barely visible to the driver and typically consisted of just road speed, fuel level and mechanical parameter indicators along with various direct mechanical means of adjusting the associated systems. Dedicated instrument panels on *dashboards* appeared from the 1930s. Typically ‘**one button one function’**, they gradually increased in complexity and sophistication in tandem with the advent of early electronics and microprocessors (Kern and Schmidt, 2010). The advent of affordable satellite navigation information and moving map displays around 20 years ago heralded more significant changes to automobile secondary controls, as did early car telephones (Harvey and Stanton, 2013).

Around the year 2000 more fundamental changes to dashboard architecture started to take place with the first **screen-based multi-modal menu-based systems** brought to the market by Lexus and BMW (Harvey and Stanton, 2013). Known as ‘*IVIS’ (In-Vehicle Information Systems)* these systems increasingly manage most of the car’s secondary and information functions (Bhise, 2011). Such a system is shown in Figure 1.4. The user interacts with a screen-based system by one or more modalities such as direct touch (for enabled touchscreens), remote indirect control (such as rotary and push controls controlling a cursor remotely), voice control, or, sometimes hand gesture using air gestures or contact gestures on touch sensitive surfaces (Harvey and Stanton, 2013). The advent of screen-based systems has permitted rapid expansion and sophistication of secondary **information and entertainment** (known as *'infotainment'*) functionality in the car (Schmidt et al, 2010; Cellario 2001) to include full mobile telephone integration, online capabilities and information transfer ('integration') between the car and its user's smartphone (Damiani et al, 2009). Infotainment offerings in cars currently include music streaming, internet searches, virtual assistants, social media, television, video, and even games (Kern and Schmidt 2009).

FIGURE 1.4 EXAMPLE OF A 2016 MODEL YEAR MULTIMODAL SCREEN BASED INFOTAINMENT SYSTEM USING A COMBINATION OF TOUCHSCREEN, 'HARD' BUTTON AND REMOTE CONTROL (TOYOTA CARS)

This change has reduced the number of 'hard' switches on the dashboard to all but the most urgent or commonplace interactions. This has freed up space in the cabin but led to greatly **increased functionality**, information provision and personalisation capacities - with premium cars routinely now having in excess of 700 features or settings (Kern and Schmidt, 2009; Graf et al, 2008). Technological innovations and information provision in cars now threatens to exceed the capabilities of their drivers, with perceptual and cognitive overload regularly occurring (Giacomin, 2012a). Functions that were once operated by familiar dashboard mechanical controls may now be embedded deep in a complex menu-based structure (Harvey and Stanton, 2013; Norman, 2005) which some have called 'a worrying trend' (Burnett and Porter, 2001). Some OEMs have had to redesign such systems to make them safer or easier to use (Lanks, 2015; Harvey and Stanton, 2013). Consumers in some markets have in recent years started to rate infotainment systems as the most **troublesome or unreliable** feature of their cars, above mechanical problems (Consumer Reports, 2014).

Most controls are already to some extent electronically mediated, therefore much modern digital driver-car interaction is arguably already a subset of **human-computer interaction** (HCI). The long predicted 'glass cockpit' (i.e. free of 'hard' controls; Walker et al, 2001) is becoming a reality (Harvey and Stanton, 2013). Compounding this, road 'feel' and noise have been steadily decreased since the 1990s to produce cars which may isolate the driver from the road and their car's mechanical functioning (Walker et al, 2006). At this time of change, an **interaction design consensus** between manufacturers has yet to emerge for the dashboard's secondary controls, in contrast to primary controls which retain relatively common designs across manufacturers and markets. Various manufacturers currently deploy markedly different interaction styles (Harvey and Stanton, 2013) in regards to secondary automation, displays, layouts and input modes. In-car information systems are perhaps deliberately becoming brand identifiers (Fleischmann, 2007). Electronics and software alone now constitute about one third of a car's total build cost (Hirsh et al 2015). In the future electronics will permit even more possible interface possibilities no longer limited by mechanical linkages. New design and measurement tools will be needed to optimise such designs (Harris et al 2005). OEMs which can engineer superior usability and reputation in their infotainment offerings may benefit from increased market share and profitability (Hirsh et al, 2015).

The semi-autonomous capability now offered on some cars' primary and secondary controls (Harvey and Stanton, 2013) threatens to change the nature of the driving task (Banks et al, 2014a) and its user experience (Meschtscherjakov et al, 2015) still further. The driver's role may at times become that of 'automation supervisor'. The problems of such piecemeal automation are well documented in other fields (see Sheridan, 2002). Fully **Self-driving cars** have been under investigation for about half a century (Saffarian et al, 2012) and may more fundamentally change drivers' understanding of what driving means (Schmidt et al 2010). No consensus appears to exist over which design metaphor or paradigm will be applied to self-driving cars and their secondary controls, although a few arrangements have been suggested (e.g. Flemisch et al 2010; Abbink et al 2012). Some anecdotal evidence suggests that self-driving and highly automated cars initially feel very unnatural to operate.

While modern roads are statistically on the whole **safer** than ever before (e.g. Young et al, 2011) there are still around 25,000 road deaths each year across EU member states (European Commission, 2015). The WHO estimates that by 2030 traffic accidents will rise to be the fifth leading cause of death in the world (Herd, 2013). Drivers aged under 25 are twice as likely to be killed as drivers as a whole, while drivers aged over 75 exhibit a similar propensity (Hancock, 1999; Rakotonirainy and Steinhardt, 2009). While secondary controls are not in themselves dangerous to use, safety may be compromised when using them at the same time as driving (Harvey and Stanton, 2013). In the US, at least 18% of injury crashes in 2010 have been described as ‘distraction affected’ (NHTSA, 2012) while others suggest driver inattention may account for 80% of car crashes (Dingus et al, 2006) and this proportion may increase further as a result of greater penetration of infotainment and mobile technology (Bowler, 2012; Riener, 2012). Secondary and infotainment controls may potentially *directly* cause an estimated 22% of accidents (Dingus et al, 2006). The use of certain in-vehicle systems while driving has been shown to decrease driving performance as much as being intoxicated at the UK legal blood alcohol limit (Wynn et al 2009).

1.2 Why use a naturalness approach?

Naturalness of interaction is a growing subject in product design perhaps because functional capacities now exceed most people’s day-to-day needs, forcing manufacturers to compete at higher emotional and aesthetic levels (Berard and Rochet-Capellan, 2015) or by designing interfaces with enhanced usability and intuitiveness (Norman, 2013). ‘Natural interaction’ is an approach which might logically appear to be associated with higher levels of usability and intuitiveness (O’Hara et al, 2013). Designing for ‘natural interaction’ might have potential to mitigate against some of the more problematic aspects (e.g. distraction, cognitive overload, perception errors) observed in modern *automotive* interfaces. Unfortunately, naturalness of interaction is inconsistently defined in general design literature (Berard and Rochet-Capellan, 2015) and almost never considered in automotive design literature. For example, some design researchers assert that naturalness should apply only to *the feeling* the user gets when interacting with an interface (Wigdor and Wixon, 2011), while others describe the interface itself as potentially *being* ‘natural’ (O’Hara et al, 2013). There is little consensus as to whether *sensory motor skill transfer*, *familiarity* or low *cognitive demand* is more important in interaction naturalness (Berard and Rochet-Capellan, 2015). Nevertheless, *some* combination of these qualities would logically appear to be beneficial in improving automotive user experience, satisfaction and even safety (Giacomin and Ramm, 2013). To achieve this, automotive naturalness first needs to be better defined, and research then needs to be carried out into what automotive design characteristics, parameters and scenarios are perceived as natural or unnatural.

1.3 Relevant definitions and research decisions

Literature was first reviewed to find appropriate definitions and classifications for all the key research terms so that terms of reference could be established for the main literature review and studies.

1.3.1 DEFINITION OF 'CAR' AND 'ORDINARY CAR'

The UK car industry representative group SMMT classifies cars as either 'Mini' (the smallest cars e.g. Volkswagen Up), 'Supermini' (e.g. Ford Fiesta), 'Lower medium' (e.g. Ford Focus), 'Upper medium' (e.g. BMW 3-series), 'Executive' (e.g. BMW 5-series), Luxury (e.g. Jaguar XJ), 'Specialist sports' (e.g. Mercedes SLK), 'Dual purpose' (e.g. Range Rover and most SUVs), or 'Multipurpose' (e.g. Citroen Picasso) [see SMMT, 2016]. Classifications may differ in other countries. The SMMT classification was considered adequate for the sampling and recruitment required by the four studies described in this thesis. In this research It was decided to focus only on the 'ordinary cars' described above, rather than commercial cars, vans, buses or heavy goods vehicles. This is because the use-cases for commercial vehicles can be rather unpredictable and extreme, often having very specific interface designs or adaptations. Commercial drivers such as taxi drivers, police drivers, and goods vehicle drivers typically receive professional driving and interface training, setting them apart from ordinary car drivers. The market for commercial type vehicles is very much smaller (around 15% of total new vehicle registrations in the UK; SMMT, 2016) and the researchers wished to focus on the majority non-professional use case for maximum impact and advancement. Electric cars and hybrid cars were included in this definition of 'ordinary' cars because they are predicted to increase in market share significantly over the coming decade (SMMT, 2016).

1.3.2 DEFINITION OF 'DRIVER' AND 'ORDINARY DRIVER'

Around 90% of the literature surveyed tended to treat car drivers as a homogenous group without sub-classifying them for analysis. Some sociological studies categorise drivers by their car use patterns or motivations but these are rarely treated as variables in studies and are inconsistent. Transport researchers such as Steg (2005) have used a three-part classification (either *"passionate", "everyday"* or *"leisure time"* drivers). Generally, the literature makes few distinctions regarding drivers' age, unless it concerns reaction-times or situational awareness of inexperienced or elderly drivers (Baron and Green, 2006). In this research an 'ordinary driver' is therefore taken to mean an able-bodied person aged between 25 and 75 regularly operating a standard non-commercial 'ordinary car' more than three days per week for social, domestic, pleasure and commuting purposes. This excludes the youngest and oldest drivers who may lack experience or have perceptual shortcomings (McGwin and Brown, 1999).

1.3.3 DEFINITION OF 'AUTOMOBILE CONTROLS'

An automobile's controls have traditionally been divided into either two or three categories: **primary and secondary** (Harvey and Stanton, 2013), or **primary, secondary and tertiary** (Kern and Schmidt, 2009). In this research the two-category classification will be used, which has a ‘primary’ and comprehensive ‘secondary’ classification. This is more common and allows brevity in writing. While primary controls have remained relatively fixed in their outward design and appearance, secondary controls have evolved more noticeably since the year 2000 (Harvey and Stanton, 2013) and currently exhibit a lack of consensus between manufacturers. As will be seen in the literature review, they became the focus of this research.

1.3.4 Definition of ‘primary’ and ‘secondary’ driving tasks

The *primary driving task* (or ‘driving’) is a complex multitask activity consisting of interactions between the driver, the car, and the environment, via the *primary driving controls* requiring the successful integration and coordination of the driver’s physical, sensory, psychomotor and cognitive skills (based on Harvey and Stanton, 2013; Kern and Schmidt, 2009). Examples of primary tasks are steering, braking and acceleration.

Secondary driving tasks are *all the other* tasks performed by the driver that are not directly related to the safe motion control, speed adjustment and hazard avoidance that characterise the primary driving task. Controlled by *secondary controls*, *secondary systems* meet the driver’s need for visibility, information, comfort, entertainment and communication and may enhance the driving experience (based on Harvey and Stanton, 2013).

- *Secondary driving* controls support safe driving for example by operating windscreen wipers, indicators and horn and keeping the driver appropriately informed.
- *Secondary comfort* controls keep the driver and occupants comfortable and alert through controlling ventilation, window opening and seat adjustment.
- *Secondary infotainment* controls access and control information and entertainment for example GPS, radio, internet and telephone.

1.3.5 Classification of secondary controls

The literature generally classifies secondary controls as to whether they are *input* or *output* devices. Referring back to Figure 1.1, Kern and Schmidt (2009) classify possible secondary control *input* types as:

(a) button
(b) button with haptic feedback
(c) discrete knob
(d) continuous knob
(e) stalk control
(f) multifunctional knobs (with three degrees of freedom/movement)
(g) slider
(h) touchscreen
(i) pedals (not usually secondary controls)
(j) thumbwheel.

Some additions, clarifications and subdivisions were made to this taxonomy (one of the very few existing) to bring it up to date and provide decision making focus for the four studies in this research. Gesture enabled surfaces and air gestures were added, as well as voice control. Tactile (surface enabled) inputs were added to the taxonomy in anticipation of future developments. Levers were added to the definition of sliders, and distinguished from them. Multifunction knobs included joystick type selector controls with various degrees of freedom. Buttons were subdivided into paired-opposite buttons (like up and down window controls) and singular unipolar buttons (like hazard warning lights) and digital clicks such as on a touchscreen where there is usually no perceptible in or out movement. Latches, which open or close a storage compartment or door, usually sprung, were also added as a category. Referring to Figure 1.2, Kern and Schmidt (2009) classify secondary control *outputs* as:

(a) analogue dial
(b) alphanumerical or digital 'dial'
(c) virtual dial
(d) indicator lamp
(e) shaped indicator lamp or icon
(f) menu display
(g) digital display.

Again, some improvements are suggested to bring this taxonomy up to date with recent developments. *Head-Up Displays* (Laser projected information onto the windscreen or 'HUD's) may be included. The classification omits non-visual modalities such as haptic feedback (by movement or vibration) and audible feedback (which may be classified as audible vocal, audible 'earcon', audible alert (Bhise, 2011). Active vibrational type alerts, normally referred to as 'haptic', were added. Furthermore, the notion of 'positional feedback' was devolved from either haptic or visual feedback, since (as will be seen) many drivers appear to judge the state of a system by the *visual position* its control is in. Finally, the terminology commonly used to describe *zones* of the car cabin is shown in Figure 1.5.

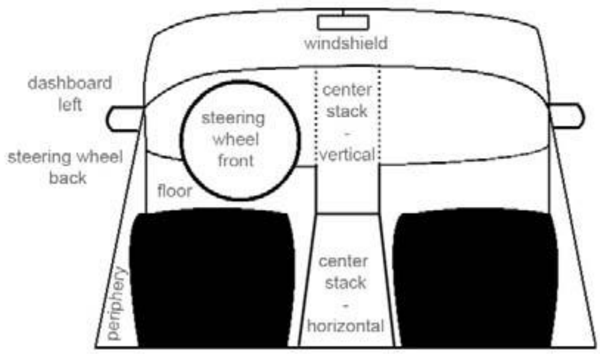


FIGURE 1.5: THE ZONES OF THE TYPICAL CAR CABIN (FROM KERN AND SCHMIDT, 2009)

1.3.6 Relevant dictionary definitions and synonyms for 'natural'

A single comprehensive definition which appeared suitable for the automotive interface design application was not identified, so standard dictionary definitions had to be sought (e.g. Harper Collins, 2017). Not every dictionary definition (for 'natural' as adjective) could apply; for example, a car control cannot be 'produced by nature' and car controls are generally not made wholly of organic materials. The full list of possible meanings of 'natural' that could possibly apply to interaction with car controls was considered to be:

1. *in accordance with human nature (e.g. "it is only natural to want to be liked";*
2. *as is normal or to be expected; ordinary or logical (e.g. "the natural course of events");*
3. *not acquired; innate, being so through innate qualities (e.g. "a natural talent");*
4. *not strange (e.g. "natural phenomena";*
5. *not constrained or affected; genuine or spontaneous e.g. "she was a natural";*
6. *following or resembling nature or life; lifelike e.g. "she had a very natural look"* (all Harper Collins, 2017).

1.4 Research aims

Consensus and excellence have clearly been achieved in many *instrumental* areas of automotive interface design over the last 100 years. At this time of rapid technology advancement however, there might be research benefit in considering what *definition of naturalness* is best suited to interaction with automobile controls, and observing and studying what *design parameters* lead to natural or unnatural perceptions. In this way, automotive naturalness might ultimately be estimated and predicted.

It is logical to focus on *secondary* controls and systems because they currently exhibit far less design consensus than automobiles' *primary* controls and systems. The outward design and layout of primary driving controls have tended to remain stable, perhaps for safety and engineering reasons, even on the latest all-electric vehicles. Secondary controls and systems, by contrast, appear to have much greater design freedom in terms of their functioning, feel, layout, appearance and input means. Secondary controls also consume an increasingly large proportion of an automobile's overall build cost and are becoming more strategically and commercially important to automobile brands (Fleischmann, 2007) while at the same time frequently contributing to negative user experiences. This research will focus on the widest use case of 'ordinary drivers' – those non-commercial drivers aged between

25 and 75 owning and driving ordinary cars regularly for domestic, commuting and leisure purposes – to make it more industrially applicable.

Consideration of what suitable naturalness definition to adopt will be made as part of the literature review, specifically in Sections 2.6, 2.7 and 2.8.

1.5 Research questions

RQ1. What are the component characteristics and dimensions of natural-feeling interaction between ordinary drivers and automobile secondary controls and in what circumstances does it tend to occur? How do these dimensions correlate with each other and with drivers' interpretations of the word 'natural'?

RQ2. What factors underlie the construct of natural-feeling interaction between ordinary drivers and automobile secondary controls?

RQ3. By measuring and rating drivers' naturalness perceptions of various interactions with their cars, can a valid reliable measurement scale for driver-car naturalness be developed that is also relevant to the future?

RQ4. How might driver-car interaction still feel natural in future more intelligent or highly automated cars?

CHAPTER 2

LITERATURE REVIEW

2.1 SEARCH STRATEGY

A literature review was undertaken to collate relevant published information so that gaps in the scientific literature could be identified, and recommendations for necessary research could be made (Risser et al., 2015). Searches were conducted in two comprehensive university libraries and online. Extensive use was made of the Bodleian Libraries in Oxford and Brunel University's Library before using the Google Scholar search engine and nine scientific and technical research databases. Dictionary synonyms were used such as 'automobile' for 'car' and these two terms are used interchangeably in the rest of the thesis.

1	**Natural/ness + Car/Automobile + Control/System (and synonyms)** • THEN Un/natural + Car • Un/natural + Control etc. • Un/natural + Dashboard etc.
2	**Natural/ness + Interaction + Driver/Car/Automobile** • THEN Un/natural + Interaction / HMI / HCI • Interaction + Driver / Car / Automobile (generally, for background theory)
3	**Naturalness +** • Design / Product Design • Controls / Systems / Machines • Mechanical / Digital / Computer / Robot Interaction / science fiction
4	**Guidelines / Measurement / Scales / Framework / Model +** • Naturalness • User Experience • Meaning / Metaphor + Design • Intuitive / Usability + Design • Instinctive Behaviour • Pleasantness + Design • Quality Of Experience • Tangible & Embodied Interaction • Human Centred Design
5	**Definition and Classification +** • Car/Automobile • Car Controls (secondary and primary) • Driver / Car User / Car Owner

TABLE 2.1: KEYWORD SEARCH STRATEGY FOR THE LITERATURE REVIEW

2.2 SEARCH THEMES

Past literature was methodically searched according to the flow chart in Table 2.1, and relevant through-references pursued on the Bodleian SOLO database and Brunel e-Library. The same search terms were inputted into the Google Scholar search engine. The keyword terms at each stage are summarised in Table 2.1. Further infographics may be seen in Appendix A. The searches began (Search 1) searching only directly related literature containing keywords ‘natural + car + controls’ (and synonyms), but this revealed very little of interest. Therefore, the search terms were widened in a stepwise way (i.e. using search terms ‘natural + car’ once ‘natural + car + control’ references had been exhausted, etc.). ‘Unnatural’ was a search term applied after ‘natural’ because the OED defined ‘unnatural/ness’ as the lack of natural/ness or semantic opposite of natural. Therefore, what makes driver-car interaction unnatural was potentially as relevant as what makes it natural. Logical synonyms were used, for example a car’s secondary controls/systems are commonly referred to as its ‘dashboard’ so this term was later used in place of ‘car controls’.

Search 2 essentially used keywords to find definitions of ‘naturalness + (automotive) interaction’. To widen this rather limited field of reference, the term ‘unnatural/ness’ was searched after the term ‘natural/ness’ Since very limited literature was found for *automotive* naturalness of interaction, all literature containing definitions of ‘naturalness + (human-machine) interaction’ from *any* discipline then had to be searched in Search 3.

Search 3 therefore concerned any type of ‘natural interaction’ potentially relevant to driver-car interaction. Compared to many human machine interactions, driving was found to span a wide range of interaction disciplines including simple mechanical interactions (e.g. a switch press or swivel air vent) digital interactions (e.g. a touchscreen) and computer interactions (HCI; e.g. multifunction menu-based systems). With the advent of semi-autonomous capabilities, it might be argued that there are already human-*robot* interactions when interacting with a car. Therefore the ‘natural interaction’ search was broadened to complete Search 3. The disciplines believed to be most relevant (i.e. revealing at least one article with keyword ‘car’; listed on Figure 2.1) were then more generally surveyed in Search 4.

Search 4 more generally surveyed the academic fields which had emerged in Search 3 such as quality of experience, human-centred design (Giacomin, 2012b) and user experience design (Helander, 2004). Operating a car is arguably a *whole-body interaction* since the body is physically moved by the car. Controls may be activated by hands, feet and voice; feedback may be detected by proprioception, haptic senses, balance, hearing, sight and vibration. Therefore, literature on tangible and embodied interaction (Dourish, 2004) was also searched in Search 4. These searches did not use keyword 'car/automobile' because that had already been done in Search 3. It sought literature *potentially* relevant to the car.

Finally, Search 5 sought classifications of all the key nouns in the research title ('drivers', 'cars' and 'car controls') which were necessary to define the terms of reference for the various studies and for sampling and recruitment. They are summarised in this thesis only very briefly for reasons of space.

Once complete, corresponding key word alerts were set up using the Google Scholar system to find any relevant new and ongoing articles, studies and patents (this being an industrial field) published over the three-year research period.

2.3 Inclusion criteria and summary of sources

Articles had to be published in English since 1960. Initially it was intended to restrict searches to peer reviewed journals but 65% of 'related work' was in books or conference papers and not peer-reviewed journals. In total 310 relevant papers, 19 reports, 23 books and 24 websites were reviewed in detail having excluded all others. A parallel survey of any possible *cultural* references to automotive naturalness was conducted from the popular motoring press by subscribing to the weekly national UK motoring magazine *Autocar* for three years, in order to gauge public and motoring journalist's opinions particularly from its editorials and road tests. These were not included in this literature review because they are generally journalistic or opinion led. Finally, four expert interviews were conducted with academics and automotive industry professionals who worked directly on interface design.

2.4 Structure of the literature review findings

The main findings of the review, where they were relevant to the research that follows, are presented below. In essence all findings below relate to at least one of the key themes 'natural interaction' 'driver-car' and 'secondary controls'. The broadest findings make reference to only one key theme (philosophies of 'natural interaction' in both the broad and the HMI/HCI sense) and are presented first (Sections 2.5 and 2.6). Next are the slightly more specific findings which target two key themes (such as 'naturalness' and 'secondary controls' (Sections 2.7 and 2.8) which make reference to the minimal number of studies which have directly addressed or referenced all three key themes at the same time. These include a summary of proposed 'natural' interventions/improvements for car secondary controls (Section 2.9), and the proposed more ‘natural’ modes of interaction for car secondary systems and their controls (Section 2.10). The absence of any peer reviewed studies on natural driver-car interaction with secondary controls made a search on ‘unnaturalness’ necessary, revealing much evidence suggesting possible *unnatural* aspects of current cars’ secondary controls (Section 2.11). This again provided an important starting point for the research and is included also because it provides a reference point for the study in Chapter 4 which asked drivers about unnatural interaction. Finally related automotive studies which used *methodologies* which showed potential for studying naturalness of driver-car interactions are summarised in Section 2.12 where that method was later considered or used in the present research. This is not strictly ‘related work’ because it did not relate to all three key themes; however it may be considered ‘related methodology’. More detailed methodological considerations are presented at the start of each study chapter (rather than in this chapter) to improve readability.

LITERATURE REVIEW FINDINGS

2.5 What is 'natural' human interaction?

This research set out to discover and measure the aspects of 'natural' interaction between humans and the secondary controls of their cars. The literature review would therefore have been incomplete without some consideration of how human beings as a species 'naturally' communicate and interact with their environment, and with each other. Some HCI research also suggests convincingly that humans, whether or not they are aware of it consciously, treat intelligent machines much as humans (e.g. Reeves and Nass, 1996). Much literature discusses humans' tendency to anthropomorphise or zoomorphise cars (e.g. Demirbilek and Sener, 2003; Windhager et al, 2008; Windhager et al, 2012; Pressnell, 1994, Marsh and Collett, 1986). Space permits only the briefest of summaries.

Natural human-human communication tends to be described as primarily verbal-auditory but studies show (e.g. Streeck et al 2011) that this is accompanied with a host of complex gestural and bodily communication cues which may convey much more information than the accompanying speech alone. A preference and need for redundancy (i.e. duplication) in communication is generally accepted as 'naturally' human (Hancock, 1999).

Humans' capacity for perception of audio, tactile, visual and environmental stimuli is well described and studied (Wickens et al, 2013) and it is sufficient to say that humans have certain strengths and weaknesses in their responses to such stimuli that are a product of our evolution. Perceptions are therefore not always completely accurate, linear or absolute, and may vary with age, gender and individual. More than any other primate humans have remarkably developed hands and capacity for both fine precision motor control and strong 'power grip' actions (Wilson, 1998) for exploring and manipulating the world. These may in some ways be considered 'natural' interactions.

2.6 What is 'natural' interaction with machines and computers?

Various frameworks and measurement scales exist for subjective interface qualities such as usability (e.g. Norman, 2013) 'quality of experience' (Möller and Raake, 2014) 'pleasurable' or 'emotional' qualities (e.g. Jordan (1998). However, no such scales, models or frameworks exist for the perceived naturalness of an interface. No accepted definition appears to exist for what naturalness might mean in the automobile or for machine interaction generally.

The notion of ‘Interaction naturalness’ has in fact been criticised for being an overly broad and sometimes ill-defined concept (Bérard and Rochet-Capellan, 2015) and this was borne out by the literature review. There is some consensus regarding its core concepts of sensory-motor skill transfer (ibid), and a feeling of learned expertise (Wigdor and Wixon, 2011). In the broader design and interaction research literature, the terms ‘naturalness’ or ‘natural interaction’ have been variously interpreted as meaning:

- richness of interaction (e.g. Jensen *et al.,* 2005) i.e. actions which are compatible with human skills and capacities and which unfold temporally, spatially and even socially, and in doing so allow building up new skills of the mind and body – as opposed to the typically ‘easy’ but ‘binary/unskilled’ human-computer interactions of the present day;
- physical/bodily interaction (e.g. Hornecker, 2011) with an emphasis on using the whole body not just the hand, or manually manipulating and moving novel ‘tangible’ input devices other than the keyboard and mouse, whose operation is more intuitively mapped to the resulting computer/machine action in three dimensions compared to relatively abstract (and usually two dimensional) keyboard/mouse/screen interactions – “*bringing the richness of physical interaction into the sterile world of digital information”* (ibid p19);
- mimicry of some measurable physical property of the natural world – whether human, animal, plant etc. (e.g. Goodman *et al.,* 2008) - i.e. a more literal interpretation of the word ‘natural’;
- mimicry of natural physics (e.g. Malizia and Bellucci, 2012; Susini et al 2012) for example referencing a familiar sound or property derived from the laws and actions of physics – specifically a *causal* property (like the skidding of car tyres to indicate a road hazard) as opposed to an arbitrary or iconic representation like a ‘chime’;
- mimicry of some real-world human action on the world (e.g. Jacob *et al.,* 2008) specifically referencing the four ways in which humans typically understand and manipulate the world – through an innate understanding of earthly physics, bodily awareness/skills, social awareness/skills and awareness/manipulation of the surroundings and wider environment;
- mimicry of human-human communication tendencies (e.g. Bickmore and Cassell, 2005, Marge et al 2010; Van Dam, 1997) for example speech, gesture, and haptic communication that is familiar from prototypical human face-to-face communication;
- similarity to human-animal interaction (Flemish et al 2012) i.e. using familiar human-animal relationships such as ‘horse and rider’ as a metaphor for some human-machine interaction in order to make an unfamiliar interactions more intuitive;

- use of established ‘cultural’ HCI interactions like the QWERTY keyboard and mouse (e.g. Malizia and Bellucci, 2012) because these artificial interaction methods are so longstanding, familiar and ubiquitous in many cultures that they have become in some sense ‘natural’;
- interaction which closely matches human mental models (e.g. Goodrich and Olsen, 2003) i.e. where “well-calibrated mental models are available, well-known sensory stimuli receive attention” invoking “well-practised response generation” (ibid) examples cited are sketching and speech; or where no learning is required (Celentano and Dubois, 2014);interaction with coherent design metaphor (Celentano and Dubois, 2014) i.e. that natural interaction always involves a metaphor in the interface design and that this metaphor is coherent, faithful and consistent in its mapping in all ways: from *source* to *target*, between its *interface* representation and its *application* representation, and the way in which both source and target are represented on the digital interface;
- interaction which feels highly familiar and instinctive as a result of skill and expertise but is solely the property of the *interaction* between human and machine – not a property of any machine or interface itself because it depends on who the operator is and that individual’s skills and experience (e.g. Wigdor and Wixon, 2011).

In general naturalness is considered a positive (desirable) quality. In the wider literature here is often an implicit assumption that the user will ‘know naturalness when they feel it’ and that it is a quality to be strived for. More recent discussions of *natural user interfaces* may be preoccupied with gestural interfaces, particularly non-contact gestures, which may have real practical limitations in cars (e.g. Norman, 2010). Indiscriminate use of the term ‘natural user interfaces’ has apparently lead to criticisms (e.g. Malizia and Bellucci, 2012, Norman, 2010; O’hara et al, 2013). In Malizia and Bellucci (2012) it is *instinctive* gestures or movements which can be considered to be natural:

> *“Unconscious movements can be considered the most natural ones, as they represent the honest signalling that happens without us thinking about it” (*p*37*).

Discussions of ‘natural’ interfaces in pre-2000 literature most often refer to *voice* interfaces in the sense of ‘natural language interaction’. Naturalness of sonic and haptic interfaces is emphasised in Whitaker et al (2008) and Susini et al (2012).

Metaphors are frequently considered important in natural interaction with machines and may be fundamental to the designer’s conceptual model and the user’s mental model (Neale and

Carroll, 1997). Metaphors enable users to experience one kind of thing (the 'target') in terms of another which they already understand (the 'source') via an implicit relationship which reduces the cognitive resources and abstraction needed. Metaphors are a fundamental part of humans' conceptual system of thought and action (Neale and Carroll, 1997).

O'hara et al (2013) summarise natural interaction as that which mimics aspects of the real world or draws on humans' existing tendencies. Similarly, Malizia and Bellucci (2012) state:

> *"By means of a natural interface, people should be able to interact with technology by employing the same gestures they employ to interact with objects in everyday life".* [p36]

Using a sensory-motor approach, Berard and Rochet-Capellan (2015) argue naturalness should be about transfer of existing sensory-motor skills and knowledge from the *real* physical world into *digital* interactions and that, crucially, this is one of the few objective ways naturalness might ever be measured:

> *"A frequently cited aspect of natural HCIs is their ability to benefit from knowledge and skills that users develop in their interaction with the real (non-digital) world. Among these skills, sensory-motor abilities are essential to operate many HCIs. This suggests that the transfer of these abilities between physical and digital interactions… could be considered as an objective measurement of the sensory-motor grounding of naturalness".* [p47]

Another attempt at a definition comes from the field of robotics and human-robot interaction:

> *"in terms of cognitive information processing, "naturalness" means that well-calibrated mental models are available, well-known sensory stimuli receive attention, and well-practiced use of short term memory is employed… naturalness… involves well-practiced response generation"* [Goodrich and Olsen, 2003, p3946].

Most in depth explorations concur that naturalness of interaction is not a property of the interface itself but of the *situated interaction* in all its various contexts (O'Hara et al, 2013; Suchman, 2007). Because interaction is only 'natural' in a specific temporal, physical, social or emotional context thus may be better understood as the *feeling the user has during such an interaction* (O'Hara et al., 2013) or how the user feels when using an interface (Wigdor and Wixon 2011) and that designers should aspire to making the interface feel like an extension of the user (Wigdor and Wixon 2011). in O'hara et al (2013) the view is that:

> *"Naturalness [in computing] is not something to be represented, but is rather an 'occasioned property' of action… produced and managed together by people in particular places – particular occasions" [p121]*

In her seminal work on office workers and copy machines, Suchman (2007) similarly showed that actions in HCI are naturally 'situated' both physically and socially. Rather than setting out with pre-set plans, humans choose what to do and how to do it in response to the opportunities of the moment (Jensen et al, 2005) – in other words meaning is made at the point of interaction, not prior to it (Boess and Kanis, 2007).

Reality based interaction (RBI) literature frequently refers to 'natural interaction' and was often uncovered by the search terms. It is considered a relevant approach and worthy of brief summary because the importance of the body in driver-car interactions has already been noted above (as compared to interactions with for example a desktop computer or household white goods). The traditional 'western' underrepresentation of bodily actions may have its origin in centuries of 'mind body dualism' i.e. that mind and body are regarded as separate entities (Jensen et al, 2005). RBI and the related fields of tangible and embodied interaction are concerned with reintroducing physical and bodily 'feel' into the 'sterile' world of digital information (Hornecker, 2011) drawing upon users' pre-existing knowledge of the everyday non-digital world (Jacob et al, 2008). Such knowledge may be drawn from users' innate sense of physics, their body and their environment and may therefore be considered 'natural'. Hornecker (2011) asserts that humans are *"living, experiencing and feeling bodies"* (p19) and that through touch, a whole host of other perceptions are available such as resistance, temperature, surface quality, softness and weight (Heim, 2008). Wilson (1998) draws together an array of neurological, anthropological and evolutionary evidence to argue, similarly, that a 'natural' human interaction is to use the *hand* to explore and learn complex skilled interactions which ultimately bring immense productivity and satisfaction. Similarly:

> *"Currently the actions required by electronic products are limited to pushing, sliding and rotating… yet humans are capable of far more complex actions"* [Jacob et al, 2008, p4]

Definitions of naturalness which centred on ease of (first) use and 'intuitiveness' were sometimes in conflict with the kind of naturalness which comes from long-term 'embodied learning' or 'muscle memory'. Hornecker (2011) believes that while touch-based interaction traditionally assumes maximum ease of use, embodied or movement interaction may advocate a more skilful learned interaction which is richer and more satisfying. An example is a musician using their favourite instrument. In a similar paradox Malizia and Bellucci (2012) argue that HCI can *become* 'natural' through familiarity: even where a gesture is not the most 'natural' (instinctive) one, it can become 'natural' following widespread adoption. They cite the example of Apple's "pinch or stretch" gesture for 'zooming' on touchscreens.

Structuralist philosophers consider that abstract concepts like naturalness and their semantics are unstable. They may and change over time and with consumer experience, culture and prevailing stereotypes (e.g. Gentner and Grudin, 1985) just as language is appropriated for a specific purpose at hand and is distinct from meaning (Hintikka, 1979). The notion of automotive naturalness can be expected to be equally unstable at this time of rapid advancement in both consumer and automotive electronics, and the adoption of semi-autonomous and electric vehicles into the mainstream consciousness.

2.7 'Naturalness' approaches relating to automobile controls

There was no agreed objective measure or physiological proxy found in the literature for measuring or estimating perceived naturalness of automobile controls. Indeed, naturalness has not generally been considered a *desirable* or even *relevant* quality in driver-car interaction until relatively recently, in particular since the growth in screen-based infotainment controls. Only one study was found directly addressing the naturalness-type perceptions of car controls. Black (1966) explicitly sought the meanings and underlying perceptions drivers attributed to their cars and (mainly) their *primary* controls. A physician and early ergonomist, Black first interviewed fifteen drivers in depth, one-to-one, about the perceptions, meanings and expectations for their driving controls. He then *hypnotised* them and asked them the same questions. He found very substantive differences in the way subjects talked about driving consciously and 'unconsciously', for example markedly different attitudes to safety, risk, and speed. Under hypnosis for example he found that:

- Drivers desired a car capable of "*taking great risks*" and were more afraid of "*getting stuck*" or "*dead ends*" than accidents.
- Drivers conceptualised the right pedal as "*power pedal*" or "*speed pedal*" but not "*throttle*"
- Drivers found the self-centring steering wheel natural and intuitive "*we expect the car to continue in a straight line*" and steering represented security and "full control".

Very few studies were found relating to naturalness type perceptions of *secondary* controls or systems. Black (1966) made occasional reference to secondary controls and based on drivers' statements under hypnosis suggested innovations in secondary controls that are arguably now commonplace, such as steering wheel secondary controls, haptic alerts, HUD displays, and Car to Car communication. He found that drivers were apparently accepting of arbitrary or even counterintuitive mappings in secondary controls such as the up/down movement of column stalks to indicate right/left, and suggested that these were quickly learned and soon became unconscious. Occasional more recent work has explored meaning and metaphor in advanced automotive secondary systems, or sought out drivers' mental models, but this work has exclusively considered only highly advanced safety systems (e.g. Kazi et al, 2007; Vadeby et al, 2011) advanced displays (e.g. Broy et al, 2012) rather than ordinary secondary, comfort and infotainment controls. One possible exception is Gellatly et al (2010) which studied drivers' naturalistic interactions with multifunctional controllers, although the conclusions were predominantly lifestyle focussed They found that drivers' everyday lives and schedules "*do not stop at the close of the [car] door*"; driving was found to be "often the least important thing going on in the car". Drivers often found in-vehicle technology complex and intimidating to learn; they expected to "extract and use previous knowledge gained from prior vehicles and the consumer electronics world" yet this transfer of knowledge and skills appeared not to be supported by luxury vehicles in 2010.

Black's study is now over 50 years old, and used methods that would not be considered ethically permissible or even scientifically valid today. Neither was he explicitly researching naturalness – but many of his research questions sought it implicitly and more so than any other study surveyed. The findings may therefore be treated as potentially having some relevance to the present research questions. From studying his writing, Black's interpretation of 'naturalness' probably tended more towards the first half of the list on page 22-23, i.e. considerations of mental models, familiarity and instinct, affordance and perhaps metaphor. Given the methods used and the era the research was conducted in, Freudian frameworks of the 'subconscious' were clearly heavily drawn upon. These presuppose there is an underlying part of the human psyche fairly consistent between individuals and always 'present' but which humans are not always aware of, and that this is intrinsically 'human' or sometimes 'shameful' but by extension 'natural' and 'honest'. It is unlikely therefore that Black's interviews sought out naturalness considerations relating to 'interaction richness' or 'whole bodily interaction' because these are more modern notions. However, he was certainly concerned with 'meaning making at the point of interaction' – to use the modern parlance. Unfortunately, he gave few details of his methods.

Interpreting Black's results from a naturalness point of view then, they could be interpreted as suggesting natural interaction with car controls sometimes goes beyond the task at hand to encompass deeper human desires, that mental models for the mechanics of car movement were either poorly understood or easily overwritten by the concept of affordance or '*what action on the world it can do for me'*, and that expectation, familiarity and full control are all important in natural interaction. Since one of Black's findings related to acceptance of apparently abstract or arbitrary mappings of switchgear to function (such as indicator column stalks, still rather in flux in 1960) it might further be concluded that naturalness is unstable but may be learned relatively easily even when the mapping is apparently illogical. Finally, the finding that drivers were subconsciously prepared to 'take great risks' despite claiming to be very mindful about safety, might help explain why public safety campaigns have not managed to eradicate dangerous driving behaviours such as checking and sending text messages while driving. It also suggests – crucially for this study – that while interview may have a role to play, methods other than narrative interview must be employed alongside it, which do not result in drivers merely giving socially acceptable 'theoretical' answers but somehow elicit, to use a cliché, their 'true feelings' and express their *actual* intentions and probably actions. Furthermore, it suggests a need for an element of ethnographic observation in an on-road setting.

2.8 'NATURALNESS' DEFINITION MOST SUITABLE FOR THIS RESEARCH

Consideration was given to all the various interpretations of interaction naturalness given in Section 2.6. Many *appeared* to provide a suitable basis for exploring and understanding driver-car naturalness, for example interpretations relating to whole body interaction and richness of interaction (given that a car evidently uses both hands and feet, as well as all the senses, and the body's sense of movement and acceleration). However, rigorous human-centred design must be conducted with empathy for the end user throughout (Giacomin, 2014) in particular anything that the end user *perceives and feels* which cannot be fully known at the outset. It was clear that none of the naturalness definitions had been intended for the driver, the automobile or its unique form of situated-yet-moving interactions. Very few naturalness definitions in the literature have been based on any direct empirical evidence, with the possible exception of the sensory motor skill transfer interpretation (e.g. Berard and Rochet-Capellan, 2015). Often these definitions appear to be 'design writing' and design theory rather loosely employed (ibid). Employing such a rigid *a priori* definition for what must be considered natural or unnatural therefore appeared self-defeating in terms of furthering knowledge. Such an approach would likely be inherently circular in nature, and would not further knowledge through discovering what *uniquely feels natural* when drivers interact with automobile controls. Such an approach would also have limited benefit for designers of improved cars, because any recommendations for more natural cars would be rather superficial and limited to the stereotypes of the particular definition used. For example, if 'natural' is simply equivalent to whole body interaction, then to make a car more natural, simply 'increase its whole-body interaction'. Indeed, it could already be estimated, or indeed even measured, to what extent automotive interactions are 'rich' 'physical' or 'bodily' but it could not be guaranteed that simply increasing 'richness' or 'bodily' aspects would *feel* more natural to the driver. As more and more naturalness literature were studied, the interpretations which focused on 'feeling natural' therefore seemed more appropriate.

Using one of the more nuanced cognitive or metaphorical interpretations of naturalness (such as coherence of design metaphor, or similarity to human-human or human-animal relationships) would again be self-referencing and likely give prejudiced or barely credible results because it would be very difficult to explore or observe such nuanced aspects other than asking drivers directly.

A more human-centred rigorous approach therefore appeared to start with the question 'what *feels* natural to drivers interacting with a car' using a sensible starting definition of 'feels natural' based on commonly understood dictionary definitions of the word natural listed in the previous chapter. Only senses of the word 'natural' which could logically be applied to the controls of a car would be considered. Such an approach could genuinely further knowledge in the field and be more reliably be used to gauge or estimate naturalness by other automotive researchers. Discussions at the end of each study could then reference all the existing interpretations of naturalness listed in Section 2.6, and explore to what extent the study's findings appear to uphold or contradict them. By focusing on the *feeling* the driver has (or imagines they *might* have) this approach would also allow consideration of future automotive interactions that have yet to become reality. Furthermore, such an approach would appear to concord with the most often cited Wigdor and Wixon (2011) interpretation of naturalness i.e. that naturalness can only be conceived of as an interaction that *feels natural* to a particular user of a particular machine in a particular scenario. This interpretation of naturalness is in fact cited more than five times more than any other with the exception of Suchman (2007 etc.) who did not explicitly consider naturalness. It would also reference that seminal work which concluded that human-machine interactions are both situated and occasioned, and that meaning is made (and perceived) at the point of interaction, not before. Such an approach to this research would of course involve the more time-consuming methodology of studying and exploring actual interactions 'in the field' (as Suchman and others did) in addition to exploration through narrative and other means.

2.9 ‘NATURALNESS’ INTERVENTIONS PROPOSED IN LITERATURE RELATING TO SECONDARY CONTROLS

There are numerous examples in the literature of interventions and innovations (tested or proposed) which claim to improve driver-car interaction on more or less ‘natural’ principles. Most describe no user input into how that intervention or innovation was arrived at and in most cases it has to be assumed it was a ‘designer’s vision’. This goes against a fundamental principle of human centred design (Giacomin 2012b). In addition, where there were any, almost all the user studies took place in medium or low fidelity driving simulators. Nevertheless, these studies may give possible insight into what might reasonably be considered by experts to be ‘natural’, in the secondary control driver-car interface. They are summarised below in list form for reasons of space.

‘Natural’ **tactile and haptic input** interventions were the largest category proposed in the literature. They include: Adding physical feel to touchscreens using inflatable or electrostatic ridges etc. (e.g. Spies et al, 2009; Kern and Pfleging, 2013); A return to “simple controls” with more tactile differentiation (Burnett and Porter (2001, p523); Mechanical controls that physically mimic their function for example a ventilation knob which itself actually blows a tiny plume of air (Feus, 2013); Using the top of the gearstick as a gesture tracking or clicking area (Riener, 2012; in fact this is now manifest in many new Mercedes-Benz cars); Improved ergonomic design and orientation of switches on the steering wheel by capturing natural hand postures and thumb movements (Takeuchi et al, 2000); Use of redundant left foot (in automatic transmission cars) for ‘zooming’ in and out in a map based navigation system (Kern and Schmidt, 2009).

Haptic outputs have been frequently proposed in terms of enhancing ‘naturalness’ of the driver experience. Many researchers consider navigation systems to be especially suitable for haptic alerts (e.g. Chen et al 2015). Approximately three times as many relevant studies on automotive haptic output interventions were found than on tactile ‘inputs’. These divide roughly (in descending order of number of references) into haptic seat information provision or warnings (e.g. Chang et al. 2011; Ho et al, 2006), haptic steering wheel warnings and information provision (e.g.Van Erp and Van Veen, 2004), haptic pedal feedback/persuasion (e.g. Birrell et al 2010, Várhelyi et al, 2002; these studies tend to be about encouraging fuel efficient driving); wearable vibrating belts or seatbelts (e.g. Ho et al, 2006). A small number of studies are concerned with adding haptic feedback to multifunction controllers or touchscreens e.g. (Pitts et al 2009). Another way of categorising these research papers is according to whether the haptic output is 'alerting' or 'informing', for which there is an approximately even split. They constitute the second largest category of references.

The third most commonly described naturalness interventions comprised **verbal/auditory inputs and outputs** – usually speech. Speech interface interventions have generally led to slightly less workload and reduced 'eyes off road' time in user testing. Improvements have been suggested such as simpler system organisation (e.g. the same commands work in the same way at every level in the hierarchy), better communication of system state (e.g. microphone on or off), more flexible allowable data entry (e.g. not constraining pace, order or format of spoken commands), choice of more 'natural' or free language for commands, not interrupting the user, and easier recovery from errors (all from Chang et al 2009). Several studies have proposed that voice is ideally suited to 'search' type functions in the car e.g. Scjmidt-Nielsen et al 2008; Weng et al 2006; Larsson and Villing 2007) i.e. to narrow down a vast number of potential options (actions) to a much smaller targeted number of actions presented on a screen based on probability. Young and Birrell (2012) suggest making auditory warnings in cars ecologically representative using 'auditory icons' such as the sound of rumble strips for lane deviation **(**see also Graham, 1999); Improved directionality of audio alerts to give richer information about hazards in '3D' is suggested in Chen et al (2015). Shifting to an auditory modality once the vehicle is moving, returning to visual and text-based interaction when stationary was proposed by Noy (1997); Adding multimodal and sonic redundancy in cars was proposed by Kun et al (2013).

Finally, there were occasional **visual, gestural or gamification** naturalness interventions proposed for secondary controls in the car. Enhanced visual displays (stereoscopic 3D or 2D with 3D metaphors) or larger visual displays have been proposed to help create more 'natural' interfaces in cars (e.g. Broy and Rumelin (2012) and Carrabine and Longhurst (2008) though many have argued the visual modality is already overloaded; 'Active speedometers' which make applicable speed limits more salient have frequently been proposed (e.g. Kumar and Kim 2005); HUDs (e.g. Liu and Wen 2004), augmented reality HUDs (e.g. Charissis et al 2011) have been proposed, as has expanding visual displays across the whole dashboard (e.g. Doshi et al 2009). A hand gesture based car infotainment interface was purported to be "more natural and intuitive" than current interfaces **(**Ohn-Bar et al 2012 p111); Easier linking of car with smart phone (e.g. de Melo et al 2009), integrating appropriate social media into the car and driving context **(**Juhlin, 2011) and gamification of fuel efficiency, driving knowledge or driving performance (e.g. Hoffman et al, 2013; Shi et al, 2012) have also been proposed and tested.

2.10 STUDIES PROPOSING 'NATURAL' MODES OF SECONDARY INTERACTION

Many examples were found in academic or expert industrial textbooks and design writing suggesting improved *modes* of interaction in future car interfaces which are described in a way consistent with enhanced naturalness. They are summarised below by *mode* in the order of most 'natural' references (greatest first).

Speech is often called *the "natural mode of communication"* for humans (e.g. Fernandez-Martinez et al., 2012) and voice interfaces have frequently been recommended where the human's hands or eyes are occupied (e.g. Cohen and Oviatt, 1995). Speech control has been seen in production cars since 1996 (Baron and Green, 2006). 'Naturalness' is often an explicit aim of future voice systems (i.e. ease of understanding, natural language not command words, mirroring human conversational tendencies and phonics, machine learning; Fernandez-Martinez et al., 2012). A meta study of all voice studies in cars before 2006 (Baron and Green, 2006) found that drivers generally drove at least as well, if not better, when using speech interfaces than with manual interfaces.

Haptic outputs are often considered to be natural or at least intuitive (e.g. Chen et al 2015). Tactile warnings have been shown to be "*inherently directional"* and "*automatically alerting*" (Ho et al, 2006 p988) because skin receptors are spread over the whole body. Many researchers consider navigation systems to be especially suitable for haptic alerts in the car (Chen et al 2015). Haptic alerts have been shown to be the most quickly perceived alert type in a car. Audio warnings are preferentially perceived over other modalities but rated subjectively more 'annoying' by drivers (Ho and Spence, 2012).

Gestural control is frequently proposed as an important new input mode in 'natural' interaction design (e.g. O'hara et al, 2013) and often recommended for cars. This has in recent years been accelerated by effective low cost gesture recognition systems becoming widespread as a result of computer gaming (Hornecker, 2011). Gestural interfaces are frequently called "natural user interfaces" (NUI) in the literature.

Driver emotional monitoring, adaption, learning and regulation are sometimes considered 'natural' because it mimics human-human interaction. Accurate recognition of emotion has been considered key to improving HCI naturalness in cars (Eyben et al, 2010). Petersson et al (2005) proposed driver monitoring (physiological as well as eye tracking, gaze monitoring, and intent estimation) in order to provide more useful warnings intuitively and unobtrusively, mimicking a co-pilot. The two main strategies for dealing with driver emotions have been described as either ‘counter-steering’ them (e.g. avoiding traffic) or adapting to them (e.g. emotional regulation or matching in-car voice with driver state (Eyben et al, 2010; Harris and Nass, 2011).

2.11 Evidence of potentially unnatural aspects of current interaction with cars’ secondary controls

Because little evidence was uncovered on *natural* interactions with automobiles it was logical to seek what aspects might feel *unnatural* to drivers. While, again, no studies directly concerned the unnaturalness of car controls, many automotive ergonomic and design research texts used the word ‘unnatural’. Evidence was often circumstantial, opinionated or of unclear origin. It cannot be described as ‘related work’ because it did not seek to explore or explain what un/naturalness means nor measure it. Some academics have suggested that the current driver user experience is poor (e.g. Norman, 2005; Schmidt et al 2010), overly complicated (e.g. Meschtscherjakov et al, 2011) distracting (e.g. Wynn et al, 2009) or disconnected (e.g. Walker et al, 2006). In summary, characteristics most mentioned in tandem with *unnaturalness* (most volume of literature first) were:

- Excessive features or unnecessary functionality, poorly integrated systems
- High complexity or workload demands
- High visual demands
- Reliability problems, inconsistencies, miscomprehensions
- Unfamiliar interactions or using different conventions to other non-car technology
- Requiring unnatural-feeling inputs (gestures, commands)
- Requiring physical gestures or postures that are uncomfortable for the body.

Offering more secondary control **features** than competitors may be seen as important in gaining market share by OEMs (Weinberg, 2008). This has led to some modern cars being described as too complicated and unnecessarily overloading drivers, causing distraction and even safety risks (e.g. Norman, 2005; Norman, 1990). Chen et al (2015) describe a typically granular, visually-dominant, poorly-integrated pattern of information systems in cars with poor coordination between them, which together might be taken as ‘unnatural’. In Meschtscherjakov et al 2011 it is argued *“Norman’s idea of the disappearing computer seems to [not] hold true… contemporary cars are often cluttered with buttons, knobs and touchscreens…. [causing] a high level of mental workload and distraction” (p5).* Studies such as these suggest such screen-based systems may feel unnatural to use when driving.

Many automotive researchers are especially critical of multifunction systems with **graphical displays** and describe them in arguably unnatural terms. Burnett and Porter (2001) call them "*a worrying trend*…*designed more for the eye than the hand*". Some go further, suggesting that the multimodal, hierarchical menus inherent to these systems are inherently unsuitable for the road (Norman, 2005) or that, by requiring the driver to search through different menus they lead to great visual or auditory distraction (e.g. Burnett and Porter 2001) increasing cognitive load (Kern and Schmidt, 2009) or visual overload (Ho et al 2006). "*For controls that were once only a button press away… drivers must now navigate through multiple hierarchical menu structures…"* (Burns et al, 2005 p1). Current automotive remote touchpad interfaces may cause increased visual demand (Sheik Nainar et al, 2016) while direct-touch touchscreens have a problematic trade-off between either a potentially straining raised hand-arm posture *or* a lower level screen below the driver's natural line of sight (Harvey and Stanton, 2013).

Backing this up empirically is evidence from ordinary drivers via the automotive press and **consumer reviews**. In 2012 the prominent US annual automotive survey Consumer Reports reported for the first time that the biggest issue reported in that year's cars were with the audio, infotainment and navigation systems (Lavring, 2012), and that this issue was the source of more complaints than engine or transmission issues (Lo and Green, 2012 p1). In 2014 infotainment was again the most complained about feature. In 2016 Consumer Reports surveyed US car infotainment by brand and found numerous usability problems regarding counterintuitive language, connectivity, slowness, crashing, over complexity, and control over/under sensitivity – many of these potentially 'unnaturalness' issues. The worst systems showed just 40% satisfaction (Consumer Reports, 2016).

Use of current **voice control** features for car secondary controls is reported to be low (Weinberg, 2012) despite such systems having been in production for 15 years or more. There is some evidence that it may be perceived as unnatural. Some automotive researchers (e.g. Broy et al 2012, Broy and Rumelin 2012) have questioned the research emphasis on voice interaction and its supposed naturalness suggesting that people have inhibitions about talking to 'machines', especially in the semi-public context of the car cabin. A metastudy of 15 papers concerning speech interfaces in cars by Baron and Green (2006) showed that using a speech interface while driving was still often worse (in terms of driver performance) than driving alone, and at that time still required the user to memorise a long list of commands, something considered unnatural (Malizia and Bellucci, 2012). OEM voice systems do not yet offer the type of non-verbal cues considered to be important in human-human communication (Lo and Green 2012). Voice interfaces tend to create high user expectations of naturalness (Hung and Gonzalez, 2013) yet have tended to be clumsy in recovering from errors (Heim, 2008). Chang et al 2009 conducted usability testing of a commercial 2009 model year VW voice control interface and found key failings in many aspects considered classically important (e.g. Norman, 2013) such as system state identification, mode errors, error recovery and system feedback. Participants failed to complete tasks by voice in up to 26% of cases and several simple tasks frequently took longer than three minutes.

Gestural interfaces are becoming more common in cars' secondary control interfaces. Norman (2010) has criticised the assumption that gesture is natural or intuitive – partly because the available actions are not visible or easily memorable and the interfaces do not promote exploration or demonstrate the possibilities available to the user. Gestures may also be dangerous in cars (Shedroff and Noessel, 2012). Gestures can also be tiring and natural only for a small subset of actions while many actions in a car will have no "natural" gesture (Shedroff and Noessel, 2012). Malizia and Belluci (2012) have further questioned the presumed naturalness of gestures arguing they are a highly specific language that has to be learned like any other: features of gestures are not consistent between people and even within individuals and few are 'instinctive'. These are all evidence of potential unnatural interaction.

Walker, Stanton and Young (2006) suggest that vehicle **feedback and** '**feel'** have been steadily reduced since the 1990s despite both having been shown to play a key role in driver safety, control and satisfaction – all potentially naturalness contributors. Loasby (1995) states: "*Modern cars… have been developed in such a way as to insulate all the occupants from the outside world as far as possible… at the expense of the driver knowing what is going on"* (p4).

2.12 Related methodology in automotive literature

A review was made of the various possible methodological approaches in the literature used for assessing an automobile's controls.

2.12.1 Quantitative approaches to assessing car controls

Car controls are typically assessed either by *analytic* methods (those used to predict system usability via simulations or early prototypes) or *empirical* studies (used to collect data on user performance under simulated or real-world conditions; Harvey and Stanton, 2013). Analytic methods in the automotive field typically consists of heuristics deployed by experts

(such as, or performance measurement testing (such as Hierarchical Task Analysis or Layout Analysis). Empirical methods include *subjective* rating scales and *objective* measures. *Empirical subjective* measures used in automotive interface measurement are predominantly the System Usability Scale (SUS) and Driving Activity Load Index (DALI) which is an automotive development of the NASA TLX subjective workload assessment tool, both conducted on representative users (i.e. ordinary drivers; Harvey and Stanton, 2013). These are of relevance to naturalness since naturalness is likely to be a subjective measure. *Empirical objective* assessments for secondary controls usually derive from secondary task 'interference' studies during simulated driving. This may be measured by lateral or longitudinal driving performance such as lane keeping or speed control, secondary task times, or secondary task performance error rates (Harvey and Stanton, 2013). These can be considered to be of little potential to naturalness measurement because they rather bluntly measure human performance in rather contrived situations rather than subjective user experience and perception in naturalistic real-world conditions (Angulo, 2007).
No automotive secondary control *measurement scales* were found in common use, but Harris et al (2005) developed a multidimensional scale to evaluate motor vehicle dynamic qualities. Two studies (Karlsson et al 2003; Aslfallah, 2008) took an architectural 'feeling' evaluation scale called the Semantic Environment Descriptive and applied it to car interiors, deriving various general perceived car cabin qualities but no questions related to interaction *with* the car's controls. Four studies (Burnett and Irune, 2009; Tractinsky et al, 2009, Wellings et al, 2008, Pitts et al, 2009) have rated drivers' quality perceptions of car switchgear and instruments. Gaspar et al (2014) studied user satisfaction of automotive audio interfaces and derived a model for 'satisfaction' based on both engineering attributes (i.e. functionality) and perceived attributes (i.e. subjective perceptions). They found switchgear 'agreeability' was most strongly related to sound, aesthetics and touch (feel).

Because of the likely sensory-motor aspects of naturalness, *physical feel* of controls may be significant. Four quantitative studies (Burnett and Irune, 2009; Tractinsky et al, 2009; Wellings et al (2008), Pitts et al (2009) rated drivers' quality perceptions of car switchgear and instruments either by 'bench testing' controls (where they are removed from a vehicle and placed on a table) or in-situ testing with branding concealed. Such studies have generally found physical feel or touch to be the dominant sense – when deprived of touch participants tend not to differentiate significantly between overall quality of different switchgears. In Wellings et al (2008) the three factors statistically found to underlie 'switch feel' were 'affect' (largely aesthetic factors), 'robustness and precision' and 'silkiness' of their action. Also of note was marked 'bimodality' in the hedonic ratings of switchgear – in other words different customers may 'love' or 'hate' the same car switchgear.

There is a large body of research concerning Kansei Engineering (e.g. Jindo and Hirasago, 1997), a method for eliciting customer's 'desired feeling outcomes' for car controls and then correlating this numerically to objective engineering parameters. Kansei has been most often used in the automotive industry, initially on the first Mazda MX5 whose continued commercial success is sometimes attributed to it. Kansei methods are highly specific, tending to relate to the design of a single button or control.

2.12.2 QUALITATIVE APPROACHES TO ASSESSING CAR CONTROLS

Qualitative approaches have the potential to 'derive fruitful explanations' and 'conceptual frameworks' (Miles and Huberman, 1985, p1) particularly about a subject of which very little is known. Meschtscherjakov et al. (2011) argue that qualitative in-situ studies are much needed in automotive interface research, noting that less than 2% of conference papers at the leading automotive interface conference between 2009 and 2011 used qualitative field studies; 45% used a simulator while 21% did no user studies at all. A small but growing body of research in the last ten years suggests that qualitative approaches have a valuable complementary role. Tonetto and Desmet (2016) studied user experience by interviewing drivers in their own cars. To counter criticism that questionnaires do not uncover 'real' feelings, they generated questionnaire items using what they termed 'natural and domain specific' language drawn from the participants themselves. Transport sociologists (e.g. Steg, 2005; Steg, 2003; Tertoolen et al, 1998) have noted that drivers sometimes appear to be less than frank about their car use and attitudes in qualitative studies. They found drivers were apparently less guarded when the aim of the research was not wholly apparent.

Published automotive interface research has rarely used ethnographic methods (Spradley, 1979) while *Contextual Inquiry* (a more rapid 'design' form of ethnography; Beyer and Holzblatt, 1997) has been used occasionally to look at drivers' interactions with their cars. Of relevance are Perterer et al (2013) concerning real world GPS use, Neureiter et al (2011) about multifunctional controllers, and Gellatly et al (2010) about how luxury car drivers use in vehicle entertainment and information systems. None concerned secondary driving controls like window wipers or ventilation but all used real cars and their subjects were their driver-owners interviewed or studied in the real world. In Neureiter et al (2010) researchers sat in the back seat of a car being driven 'naturally' by its owner (their words) on a public road, to observe drivers' interactions with multifunctional rotary knobs in an in-car Contextual inquiry. Gellatly et al (2010) used Contextual Inquiry to focus on how luxury car drivers use in vehicle entertainment and information systems in the real world by placing *two* observers in drivers' own cars – one in the passenger seat guiding the inquiry and observing, and a second silent observer in the back seat observing and recording hand location and glance behaviour. Within 24 hours both researchers conducted an interpretation session with a wider group of researchers and designers to capture design insight. Laurier and Philo (1998) used Contextual Inquiry and observation to explore how people communicate while driving, in naturalistic settings. Their method was ethnographic participant-observation and they undertook many journeys over a period of months, becoming integrated and trusted observers. Driver-car interactions are usually silent, private and even unconsciously executed (Dogan et al, 2011) limiting the insight from silent observations without questioning or ‘probing’. This research offered an opportunity to further develop such ethnographic and human centred design research methods to the car cabin, and offer a methodological contribution.

2.13 Summary conclusions

In summary it may be concluded from the literature review that:

1. There is a clear research gap around the *basic understanding* of what ordinary drivers perceive to be natural or unnatural when interacting with their automobiles' secondary controls. The only directly related work is now over 50 years old. Although some interaction design researchers have explored the meaning of naturalness, no suitable definition or framework could be identified for the automotive application. An interim definition of naturalness will be required as a starting point (see Section 2.14).
2. There is evidence of usability, satisfaction and perhaps safety problems with existing car secondary controls, which might be addressed by adopting a *naturalness* approach. Furthermore, at a time of rapid dashboard evolution and innovation with novel secondary features, interaction modes and forms of automation currently being proposed, no clear design consensus or naturalness philosophy appears to exist for their interface design. There is evidence that some of these innovations may be perceived as *unnatural*.
3. There is a second research gap around what *design parameters* contribute to any natural or unnatural feeling when interacting with car secondary controls. No relevant measurement scales or models were found in the literature. Various interventions relating to secondary control design, input modes and output modes have been *proposed* as 'natural', but human-centred qualitative approaches do not appear to have been adopted; user input appears to have been scarce or obtained from simulators, results are often contradictory. A 'first approach' exploratory study is therefore required.
4. Operating a car and its controls is a uniquely *situated*, *occasioned*, mobile, real-world, whole body interaction. The unique dynamic setting of the car cabin, the unpredictability of its interaction scenarios, and the safety demands of the road, must all be considered. Study of natural interaction must also consider meanings, instincts, expectations, mental models, associations, metaphors and lived experiences as well as functional aspects. Naturalness is unlikely to be fully understood using quantitative 'human performance measurement' approaches. A qualitative human-centred approach is indicated. Exploration of naturalness should be grounded in real-world cars using real car controls in real-world settings wherever possible, and not in simulators.
5. An *industrially applicable measurement scale* for naturalness might allow more efficient assessment and improvement of critical high-cost secondary systems. OEMs who apply this naturalness measurement approach to the redesign of unpopular 'brand identifying' infotainment systems might gain reputational and commercial advantage.

2.14 WORKING (INTERIM) DEFINITION OF DRIVER-CAR NATURALNESS

In advance of discovering what naturalness of secondary control interaction means to ordinary drivers, some interim (working) definition was required in order to define objectives for the initial studies and devise questions for early interview schedules. The definition needed to be open-ended enough not to prejudice participants towards 'circular' responses to 'self-fulfilling' questions, yet sufficiently defined so participants answered the research question rather than some other question. Inevitably this would be an iterative process, with more knowledge and insight gained after each study. Based on the findings of the literature review and the small amount of relevant research, a logical starting point appeared to be:

1. **Naturalness of interaction** is presumed to be a largely *subjective* property pertaining to the *perceptions and mental operations of the human being*, but which may be triggered, enhanced or diminished by specific physical, mechanical, perceptual, relational or cognitive characteristics of the object or system which the person is interacting with, some of which *may* potentially be *measurable* characteristics.

2. **Natural driver-car interaction** must consider the full spectrum of *instrumental, symbolic and affective* interactions between driver and car. Natural or unnatural interaction is likely to be adequately described and understood simply as *interaction that 'feels' natural or unnatural to that driver during use* and this must be the starting point for the research.

3. **Natural interaction with secondary car controls** similarly must apply to the whole interaction between 'driver and control and system', rather than being a property of any single control, interface, system or dashboard. Again, it is likely to be a largely subjective 'person-centred' metric rather than an 'object centred metric'.

4. **Drivers' interpretations of 'natural-feeling' interaction with secondary car controls** *might* reasonably comprise elements of the following: *Physical/mechanical aspects***:** involving sensory-motor skill transfer; involving multisensory, multimodal, direct or whole body interaction; adherence to physical properties of the natural world; *Perceptual aspects***:** familiarity, intuitiveness, instinctiveness, simplicity, connectedness, directness; *Interactional/cognitive aspects:* meeting of interaction expectations, stereotypes and logic; easiness; mimicry of human-human communication tendencies, using innate skills.

These tentative definitions were used to help compose the interview schedules for the first study described in the following chapter (Chapter 3)

CHAPTER 3

A CONTEXTUAL INQUIRY INTERVIEWING DRIVERS INSIDE THEIR CARS

3.1 AIM

This first study needed to discover and explore as many as possible characteristic components of naturalness between ordinary drivers and the controls of a standard car. A qualitative enquiry is the normal starting point for such exploratory work (Patton, 1990) where the first stage is to obtain a 'longlist' of findings without them necessarily being representative or generalisable. By the same logic, an inductive approach (*Grounded Theory*; Glaser and Strauss, 2009) was also indicated in data analysis because of the lack of any suitable frameworks about naturalness. The study's objectives were to explore:

- driver's most *salient* and *familiar* interactions with the car generally (including with primary controls, for overall context) and their general expectations from their cars;
- drivers' perceptions of what *natural-feeling* interaction means with cars generally;
- natural and unnatural aspects of interactions with car *secondary controls* (based on the working definition of naturalness) including any *expectations* drivers have.

3.2 RESEARCH METHODS

Naturalistic context (or *ecological validity*; Patton, 1990) was considered to be highly important because the research question concerned what were presumably nuanced characteristics of a *situated* (contextually dependent) and perhaps *occasioned* (scenario dependent) quality ('interaction naturalness'). If, as the literature review had suggested, 'natural interaction' essentially means 'natural-feeling interaction' then it would be illogical to conduct this research in any setting which drivers may perceive to be unnatural, unfamiliar or artificial. This ruled out driving simulators (especially computer desktop simulations). The literature suggested a small number of feasible contextually faithful exploratory methods, and these are outlined below.

Ethnography, the scientific description of peoples and cultures 'at work' with their associated practices, customs and habits (Wolcott, 1999, Dourish, 2004) appeared to be a logical starting point for this exploratory study. A central tenet of ethnography is to observe people doing everyday activities in ecologically valid settings, rather than in constructed or laboratory situations (Helander, 2004). A pilot study involving three drivers was therefore conducted, silently observing them while they drove scheduled journeys. However, such strict ‘participant observation’ type ethnography was found to be of limited use at this exploratory stage because driving and operating secondary controls appeared to be composed of multiple private, silent, and difficult to interpret interactions between a single human and car. Commands and feedback were rarely vocalised, audible nor shared; nor are such interactions apparently always executed fully consciously (Dogan et al, 2011). Therefore, observational ethnography was considered to be more suited to a later study (described in Chapter 6) after a framework of basic understanding had been obtained.

Contextual Inquiry **(**Beyer and Holtzblatt 1997) is a less passive and more rapid form of ethnography which is sometimes used in the design of products. It mixes focussed semi structured interview, impromptu ‘probing’ questions and observation (Salvador et al 1999). As with ethnography, instructions specify that inquiry takes place ‘where the action happens’ (Beyer and Holtzblatt, 1997 p 58), specifically in the normal ‘work’ environment. Suchman (2007) also concluded that human machine interaction needs to be observed in the *situation* where it occurs (Boess and Kanis, 2007). Beyer and Holtzblatt (1997) created an analysis process which they called ‘affinity diagramming’. Contextual Inquiry has occasionally been used within the automobile (e.g. Meschtscherjakov et al, 2011; Gellatly et al, 2010).

Ethnographically purposed interview questions (Spradley, 1979) has been used to actively explore ‘habitual’ interactions which are hard to understand. In a study of fuel efficient driving practices by London taxi drivers, ethnographic interviews were first conducted inside moving taxicabs to understand drivers’ world view, community, behaviour and physical habitat (Rowson and Young, 2011). A further human centred adaption of the ethnographic interview method is to involve the users in the development of the interview questions, for example by using word association type activity (Giacomin, 2012a). This can avoid prejudicing interviewees with the researcher’s *a priori* expectations of the research. Knapper and Cropley (1980) explored the interpersonal aspects of driving qualitatively in this way while maintaining rigor. They initially conducted 13 ‘expert’ unstructured interviews with key stakeholders and randomly selected members of the public, simply asking them to talk freely about driving and noting any themes of relevance that arose. Thirty members of the public were then selected at random and again asked to talk freely within an interview schedule based around the themes drawn from the stakeholder interviews. Finally, a series of questionnaires were compiled consisting of statements and scenarios with agreement scales, based on the common themes obtained from those previous stages.

3.3 Key methodological decisions

A Contextual Inquiry conducted inside ordinary drivers' own cars, using questions derived from a series of pilot studies with such drivers, and phrased in accordance with the ethnographic principles of Spradley (1979). was considered to be the logical first approach. The ethical and safety implications of a moving on-road study with intensive questioning were considered too onerous for this first study. As a compromise, it was decided to conduct all interviews in parked cars. A semi-structured interview format (Patton, 1990) was considered to be an acceptable compromise between capturing the (as yet unknown) personal perceptions while still maintaining some degree of consistency between subjects. It also allowed unanticipated findings to arise and be recorded, which is important at an exploratory stage when the question schedule may not necessarily yet be 'optimal' (Knapper and Cropley, 1980). At this initial stage it was felt necessary to widen some of the initial questions beyond mere secondary controls, to include drivers' interactions with the *whole* car (including primary driving controls) in order to understand the overall context in this initial study. Perceptions of secondary interactions could not be assumed to be entirely separate from perceptions of primary interactions. This is in line with ethnography practice such as Spradley (1979) where the subject's experience of the overall context is explored as well as specific phenomena of interest.

In enquiring about drivers' perceptions of various secondary controls and systems, it appeared logical that the driver should already be familiar with those controls. As well as talking about the controls in their own cars, it was additionally decided that drivers should *operate* some of those controls at least once during the interview (Beyer and Holtzblatt, 1997). Participants were assumed to be close to the familiar sensations and perceptions of operating a car when seated in their driving seat and operating a control. It was hoped this would elicit truthful and realistic responses based on past experiences, expectations and instincts, and their likely perceptions and behaviour in the future.

3.4 Objective

The objective of the study was to ask a sample of drivers (representing all six common car types (from SMMT, 2016) and three main driver profiles (from Steg, 2005) to talk relatively freely in a structured one-on-one Contextual Inquiry interview (Beyer and Holtzblatt, 1997), inside their own cars. The interview questions would be shortlisted and developed in pilot phases using domain specific language (Tonetto and Desmet, 2016) derived from ordinary drivers. The questions would be structured like an ethnographic interview (Spradley, 1979).

3.5 STUDY DESIGN

The pilot stages were designed to permit impromptu ‘probing’ questions and further information and clarifications, with later stages being more fixed and consistent between subjects. The questions themselves were devised using the systematic iterative method used in Knapper and Cropley (1980). Like that study, the evolution and selection of the questions formed part of the study itself (this being part of the ‘method’ and therefore described in that section). Two pilot studies were designed to develop the questions using 12 participant volunteers, all drivers aged 25-75 from Brunel University. In order to capture as many relevant perceptions as possible, each research question was addressed in a variety of ways in order to extract the maximum amount of data possible. Following recommendations from Steg et al (2001) the purpose of the research was kept unclear until the end of the interviews, so as not to prejudice answers (Krefting, 1991). Therefore interviewees were not prejudiced directly with the word ‘natural’ in the initial two-thirds of the interview (instead using synonyms such as ‘easy’, ‘intuitive’ or ‘ideal’. All questions were ‘open’ to achieve richer narratives. Questions were aimed at exploring *actual* past experiences, feelings and meanings, to discourage socially-mediated ‘theoretical ‘responses (Cresswell, 2012). As far as possible common cognitive biases (Bless et al, 2004) were identified and controlled for in the questions, analysis and interpretation. Cognitive biases are flaws in judgment that arise from errors of memory, social attribution, and miscalculations (Dvorsky, 2013). They include *question order bias, confirmation bias, leading question bias,* and *social desirability bias*, all of which may affect interview type studies. Only once the two pilot studies were complete and fully analysed, and the potential question list negotiated with a senior academic, would the final questions be devised.

3.6 SAMPLING AND RECRUITMENT

Qualitative research explores subjective experience; validity and reliability requirements are different to those in quantitative research (Burgess et al, 2013). Sample size and participant selection are less critical because results are not intended to be generalised to the population as a whole (Morse and Field, 1995) and data can be considered ‘saturated’ simply when no new verbal ‘codes’ arise from subsequent participants (Mason, 2010). A trade off exists between the depth (and length) of interview and the number of participants used (Cresswell, 2012) given finite resources. Smaller sample sizes combined with in-depth interviews is considered a valid exploratory strategy (Patton, 1990). It was decided to conduct a relatively small number of ‘deep’ (i.e. long) interviews using the ‘n=12+3’ sample size principle (Baker and Edwards, 2012) whereby 12 interviews are conducted and fully analysed with its themes formulated into a framework, as if the study were complete. Then, three more interviews are subsequently conducted in exactly the same way simply to check if any *new* codes or themes have arisen. It has been suggested that this method often demonstrates saturation with no new codes or themes found beyond the 12 initial interviews (Guest et al, 2006).

A purposive maximum variation sampling strategy was used (Patton, 1990). All interviewees were recruited through adverts in social and professional networks, and two car clubs, in Oxford and Uxbridge, asking for details of car type, car use patterns and driver age. They were compared to the classifications of car types and driver types described in Sections 1.3.1 and 1.3.2 above, so that at least one interviewee was allocated to each of the nine categories. A ‘hybrid’ car driver and a ‘city car’ owner were included to represent possible future ownership patterns (SMMT, 2016). Roughly equal numbers of men and women were sought. No payment was offered. Screening eliminated drivers aged under 25 and over 75 because of possible perceptual limitations or inexperience (McGwin and Brown, 1999). Participants were required to own their own car or have regular use of the same car at least twice a week to fulfil the requirement of the test cars and their controls being ‘familiar’ (Steg, 2003). The Brunel University Research Ethics Committee approval letter is included in Appendix B.

3.7 METHOD

For the first pilot exercise, four drivers were asked to speak freely on the subject of the main driver-car interaction themes found in the literature review. These interviews were immediately fully transcribed and subjected to thematic analysis (See section 3.8). Key words and themes apparently relating to the research question were noted. Thirty-two possible interview questions were then conceived with reference to the ethnographic interview questioning styles of Spradley (1979) and Osgood (1957) as far as possible using the key words and themes expressed by the participants themselves.

A second pilot study of six drivers then took place using the 32-question interview schedule from above, and transcripts were analysed according to the potential each question demonstrated for answering the research question fully and without prejudice. This was done by discussing the transcripts with a senior automotive researcher (the academic first supervisor, who had 10 years’ automotive human factors experience and 15 years’ academic experience). Twenty-nine final questions were agreed by negotiation following the ‘debugging’ and rewriting process which essentially removed any apparently misleading, misinterpreted or unproductive questions and deconstructed compound questions into smaller constituents. The 29-question interview schedule is presented in Appendix C. The initial section of questions probed whole-car related expectations and secondary control/feature saliency. Central questions enquired about natural-feeling interaction in several ways and presented five short scenarios for drivers to visualise and reflect upon, so that they could draw upon a range of previous driving experiences and consider the concurrent primary and secondary interactions. Participants were asked to close their eyes and picture five relatively common scenarios (scenarios that most drivers would have experienced recently, such as waiting at traffic lights). More probing questions using the word 'natural' and 'future focused' visualisations were reserved until the end, when participants had appeared to be most relaxed and frank. These encouraged drivers to visualise interacting with an imagined natural-feeling car in the year 2030.

All 15 final interviews took place in interviewees' own parked cars. An introductory statement was read out and a consent form presented for signature. Biographical data was taken (age group in five-year brackets, car type, annual mileage, home town and job title) to give context to quotations used in the analysis and results, and to help explain any patterns in the data. The participant, seated in the driving seat, was instructed to safely place hands and feet in their usual driving positions. The 29 questions were then asked with the researcher sitting in the front passenger seat, using an audio recording device. No further explanation of the questions was allowed, to maximise consistency between participants. A sample of raw data is shown in Figure 3.1. Full transcripts (very large files) are available on request.

3.8 Data analysis

The aim of the data analysis was to find driver-car naturalness themes both *explicitly* (through analysing the responses which addressed the research question more directly) and *implicitly* (considering all other responses using a measured degree of interpretation) for any possible phenomena, perceptions, feelings, and expectations *apparently* related to naturalness.

3.8.1 Thematic Analysis and Content Analysis

Thematic Analysis (TA) is a method for identifying, analysing and reporting repeated patterns of meaning (themes) within data, and organising and describing datasets in rich detail (Braun and Clarke, 2006). TA has been successfully used in qualitative car studies before (e.g. Burgess et al, 2013) and provides "much more than… a simplistic description of data or a counting game" (Elo and Kyngas, 2007) effectively 'telling the story of the data" and interpreting it (Braun and Clark, 2006) providing an ‘illuminating description’ of the phenomena of interest (Smith et al, 2011). TA was used inductively – that is coded without trying to fit the data into a pre-existing framework. Well-established stepwise guidelines were followed from Braun and Clarke (2006) to maximise rigor and reduce subjectivity. These stages may be summarised as: familiarisation with the data, coding semantics in the data, searching for themes, reviewing themes, defining and naming themes, and writing up.

Braun and Clarke (2006) argue that a common error in TA is to conceptualise themes as ‘residing in’ the data. Inevitably the researcher plays a far more active role in identifying the themes and patterns that are of interest to themselves and their readers – involving active selection, editing and selective deployment (ibid). TA must be made more rigorous therefore by using a panel of independent ‘coders’ to corroborate or challenge the lead researcher’s findings (Krefting, 1991; see Section 3.8.3). To assist with this, the interpretation of the lead researcher was kept transparent throughout the analysis by embedding direct quotations from participants into the descriptions of the themes themselves. In this way the independent researchers could follow the logic of the main researcher, agreeing or disagreeing with the interpretation and theme development and wording. Other measures taken to reduce bias and subjectivity were full verbatim transcription (rather than note taking), comparing codes back to the whole dataset to see that they do not contradict the wider data, and eliminating ‘outlier’ codes which were only expressed by a single participant (Braun and Clarke, 2006).

Content Analysis (CA) was also used to ‘distil’ keywords and themes relating to driver-car naturalness into fewer content-related categories (Elo and Kyngas, 2008) to make patterns in the data clear. Basic word and theme counts were performed to estimate the saliency of the various car controls mentioned and estimate the strength of naturalness perceptions. Computer programs (such as Nvivo) may be used to assist in TA by organising the data but do not yet have the processing power to understand the words and themes, which remains the task of a skilled researcher (Morse and Field, 1995). Manual analysis (using paper transcripts) was chosen because of enhanced recall and pattern recognition that can derive from handwritten colour coding. A matrix framework (Smith et al, 2011) was created on a bespoke spreadsheet to summarise the findings and help identify patterns from the CA keyword and theme counting.

Imagine you could communicate with your car very naturally in the future. What would that feel like? Can you give me examples of how you would drive that imaginary car?

Um that's really difficult because you've got very much ingrained notions of what 'to drive' means. If you drop the word 'drive' and say 'travel in'… but would you feel like a driver or would you feel like a passenger.. the difference in that. I can imagine systems working better not having to make something happen it would just know that it was the appropriate time to do something, and would that decrease cognitive load? To help you driving? Probably not you'd be listening to some nonsense on the radio instead, or probably watch TV, play games whilst you're driving that would be cool [laughs].

So it would know ahead of time what you were thinking?

I think it would be able to yeah. (a) it would be able to understand but (b) it would look at the externalities outside the car and do things so you wouldn't need to then necessarily do things perhaps it starts to rain, and the wipers would come on and also it might change the ambient temperature to a way which you would want to do normally.

Any what would it mean to you to have a car that did that?

Um I think it would be a statement of luxury in one hand, your mates would be very impressed by those kind of… functions but you would normalize to it pretty fast as well, and I think you would always want that manual override option, which could be a voice interaction you just talk to you car and um yeah, not this time, and say something else, there would have to be some sort of way of interfacing with it to regain control if you wanted to.

FIGURE 3.1 EXAMPLE OF THE RAW DATA FROM AN INTERVIEW (FINAL TWO QUESTIONS)

	A	B	K	L	M	N	O	P
1	CURRENT OVERALL EXPECTATION	relavent questions	nick	sharon	john	jim	absolute	sum tot inc re
2	What do drivers expect from car NOW, other	4, 1,2, 3,11 (expect	A to B, radio listening	utility x3	necessary evil. Folding mirrors. (mus	car 'looks after itself' (potency?)		
3	mentions reliable	4, 1, 2 ,3, 11	yes, and to start	yes, start, reliability x2	utter reliability. To start quickly whe	yes, surety of arrival/ avoid embarass	10.00	16.00
4	mentions safe	4, 1, 2 ,3, 11	no, only elsewhere	safe x2, safety.	no	no	4.00	9.50
5	mentions comfort/ease/min fuss/haven	4, 1, 2 ,3, 11	yes - seat, warm, comfort, cabin	warmth	haven, isolation, solitude	yes - esp of bigger cars	8.00	18.00
6	mentions fun		likes engine sound	fun seems to = luxury	only elsewhere	no, only elsewhere	4.00	4.00
7	feeling in full control most of time (arose, not	17	probably (going to work)	yes.	on open road	yes	5.00	5.00
8	feeling in full control is…	17 and elsewhere	extension of me', feels at home,	a good feeling, responsive, conven	pleasure, sated, luck, mech perfectio	comfortable, enjoyable x3 (fast?)	sense of flow, mastery anc	
9	What are the (unnatural) (unexpected) (loss ctrl) feelings about driving and cars	18, 19 (Bad), 20, an	scary	not nice, frustrating x2, embarrase	messing yourself up, powerlessness,	frustrating x2, annoying, scary, dangerous, hairy (impotence)		
10	SALIENCE OF INPUT SYSTEMS AND FEEDBACKS		ARNED?				ABS	TOT
11	What systems are the most salient in driving - NOT 4 BELOW -NOT FEEDBACK	OTHER	demist x2, seat pos, ignition, eco	autopilot.	folding mirrors. Manual gears centra	Engine MI.	GPS	3.00
12	mentions steering	1, 11, 12-16 (sys on	SW	elsewhere	general steering at speed	hands on wheel'x2 ; steering	9.00	19.00
13	mentions speed or accel, maneuvrability	1, 11, 12-16 (sys on	yes x2 (zoom, accel), being ready	agile & responsive, small, power fo	(general speed control) preparednes	responsive', ready to go	11.00	27.00
14	mentions transmission any	1, 11, 12-16 (sys on	gears, put it into reverse, likes to	gears (use of) elsewhere	yes, central (auto a joy)	automatic x3. hand naturally on leve	10.00	20.50
15	mentions safety generally	1, 11, 12-16 (sys on	no	elsewhere	no	no	6.50	7.50
16	mentions stability or handling, tyres	1, 11, 12-16 (sys on	handing (at speed)	no	(general handling)	no	5.50	7.50
17	mentions comfoprt or ease	1, 11, 12-16 (sys on	yes x3, seat pos, climate, int tem	climate (warmth) elsewhere, quiet	climate ('invariably)	trying to get temperature right	9.00	17.00
18	mentions music/radio	1, 11, 12-16 (sys on	yes x2 - sanctuary	radio or MP3/CD - innter world	radio	radio x2, changes channels	9.50	19.00
19	mentions visibility	1, 11, 12-16 (sys on	yes x5 (mirrors, demist), winscre	no	demist, mirrors x3, park assist	watches cyclists. Checks mirrors.	8.00	24.00
20	mentions brakes and handbrake	1, 11, 12-16 (sys on	brakes, Park function in auto	stops when I need it to	elsewhere - gives security, forgives	yes, also brakes elsewhere	9.00	16.00
21	What are the most salient feedbacks from the car	2, 5, 6	general functioning, gear positio	general functioning, responsivenes	GSI (too simple!) Stop Start (too brut	loss of power = problem	ABS	TOT
22	mentions car gauges - analog ie not on/off	2, 5, 6	fuel, range, revs	temperature	revs, speedo, MPG x2 (game?) alerts	temp, revs (old/problem cars), MPG	10.00	38.00
23	mentions graphical display screen/GPS	2, 5, 6	no	from someone else's car yes.	no	console messages (text) eg oil top up	4.00	15.50
24	mentions discrete warnings binary	2, 5, 6	no	oil light	washer fluid, boot close, seatbelt,GS	on console (monitors), plus audible,	7.00	18.00
25	mentions steering feel	2, 5, 6	no	SW feels direct.	elsewhere	no	6.50	17.50
26	mentions radio or phone as feedback	2, 5, 6	no	no	no	no	1.00	2.00
27	mentions pedal feel	2, 5, 6	no	yes, accelerator feedback?	no	no	4.00	5.00
28	mentions mech sound or vibration	2, 5, 6	yes, when runs well. Responsive	sound genersl.	engine, suspension	general feel/sound x2 - problem alert	11.00	31.00
29	What are the most perceived/salient body parts in interaction	6, 12-16 if body par	eyes, hand	feet x2, hands	eyes, whole body forces sensing	eyes and ears	feet	hands
30			Unless auto				4.00	5.00
31	Attitude to primary physical-mechanical controls	26, any esp 23	luxurious feel, extension of self	autopilot. Perh extension of self. R	discipline. Loves physical controls - d	strong, direct, reassuring, if autopilot	9.50	5.00
32	RELATIONSHIP, INTELLIGENCE, DIALOGUE						important to	hands&arms
33	How do drivers relate to their car now emotionally/humanly? What do they treat it as?	3, 4 & any gen	utility, adjusts controls, gets it to	purely functional. Not attached. Lo	necessary evil. No personality (mode	modern cars no character. Fond. Shape seems to engender feeli		
34	association/metaphor	3, 4 & any gen	kitchen oven	bike - way of getting round	none. Unique object.no soul.	no (other classic - cat) Teutonic, ordered, looks after itself.		
35	What is the most intelligent part of the car?	10	engine and auto trans	head unit display (GPS, radio, GSI,	stop/start	ECU ('little')		

FIGURE 3.2 EXAMPLE SCREENSHOT OF THE CONTENT ANALYSIS MATRIX

3.8.2 ANALYSIS PROCEDURE

The full process for the fifteen interviews (i.e. 12+3) was as follows:

1. The interviews were transcribed in full including expressions and pauses, then read three times to identify semantics and meanings possibly related to the research question. The question borne in mind was ‘What are they really saying?’ (Morse and Field, 1995). After comparing transcripts person-by-person, responses were compared question-by-question (bi-directional analysis).

2. The transcripts were then trial coded at a basic semantic level, and codes combined where there was a dictionary simile. Possible higher order (interpretative) naturalness themes, patterns and potential groupings for the codes were noted in the margins. Some responses were coded from a single-word unit of content measurement (where the question was primarily about eliciting single systems or semantics – e.g. “the steering wheel”). More complex descriptions were coded thematically as the smallest measurable unit (e.g. the theme of “natural-feeling cars being proactive”).

3. Each code and theme was checked against the dual criteria of *internal homogeneity* (whereby data within a theme are meaningfully coherent and the same) and *external heterogeneity* (whereby each theme is logically distinct (Graneheim and Lundman, 2004). Some codes were reworded.

4. A fresh transcript of every interview was then fully coded with these keywords for individual codes and various colours of highlighter pen for the common themes i.e. sharing the same meaning (Cavanagh, 1997) to make patterns clear. Additional notes were made in margins of any potential overarching patterns and interpretations.

5. A matrix framework approach (Smith et al, 2011) was then used to draw all the findings together and do the basic content analysis (word and theme counts). A bespoke database was created in Microsoft Excel. The database made word counts, saliency estimation and theme strength estimation relatively simple.

6. Themes not shared by at least 30% of participants were ignored (Braun and Clarke, 2006). Codes were then helpfully ‘named’. Five codes were reworded as a result of the independent code checking process described below.

7. Finally three further interviews were conducted to check that no new codes arose. They were analysed using the same methods. No new codes were discovered. One theme was broadened as a result of the independent code checking process detailed below.

	Theme 10 VOCAL INFORMATION EXCHANGE FEELS NATURAL	
10. 1	The dashboard would say 'good morning' and 'where are we going to today' and 'what time do you hope to arrive' and you'd say back to it 'we're going to Holborn' or 'we're going to Milton Keynes' and it would say 'certainly' 'sit back, relax – the traffic is very bad on the route you normally go, we are going therefore THIS way. It's going to take you at least two and a half hours. Do you still want to go?' M 65-70 Premium it'll probably be keyless and voice activated just in terms of getting in and saying 'Awright, off we go!' M 40-45 Premium OK I think it would all be electronic, you just… the door would open for you, you just sit down, it would tell you 'OK I'm ready to give you an update' 'everything's OK' 'you've got enough petrol for this…' It would tell you all the… and you'd communicate – tell it where you want to go, and it would say which way would you like to go there, a bit like a GPS kind of, I guess you just then have an option [laughs] this is crazy – for IT just to drive you, or if you wanted to say 'Hey! I want to drive today let me drive'. F 45-50 Premium Yeah I imagine sitting in the car and 'Hello James, welcome back, where are we going today" and I would say "work, unfortunately" M 30-35 Premium	*Conversational style feels natural – building up intelligent understanding through improvised dialogue and turntaking.* *Car asks intelligent questions and responds intelligently*

FIGURE 3.3 EXAMPLE OF THE CODING SHEETS FOR THE THEME 'VOCAL INFORMATION EXCHANGE'

3.8.3 INDEPENDENT CODE CHECKING PROCESS

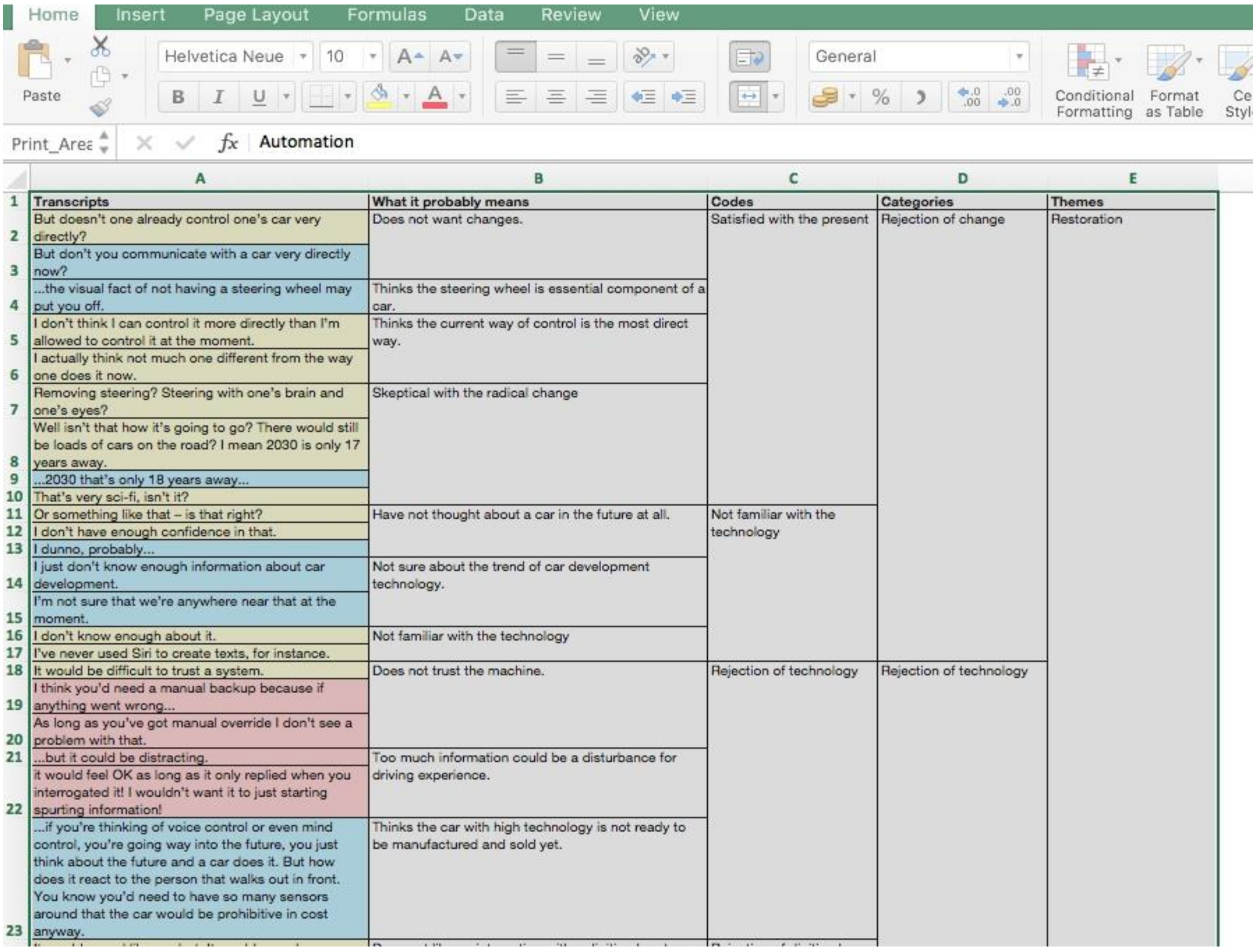
Initially, the logic and hierarchy of all the codes and themes was discussed with an independent psychology researcher from another university and with the lead academic research supervisor, an experienced automotive ergonomist and professor of human centred design. The independent psychology researcher was selected because of his master's level skills in well-being research (conducted using a variety of standardised instruments) and knowledge of social psychology and narrative methods. Minor amendments were made to some of the codes to maintain internal and external homogeneity (Patton, 1990).

A 'blind' TA code checking process was devised involving sending three randomly selected transcripts to three independent doctoral researchers within the same department but with no knowledge of the project's findings. One was a researcher of iconicity in automobile recognition, one was a researcher of vehicle seat vibration perception, the other was a social media researcher. No hard and fast rules exist for checking but the ratio of three external researchers to 15 interviews was felt to be adequate. Each was sent a simplified coding schedule with the hardcopy transcripts. Each code had a number attached. The checkers were asked to write down the number of the code onto the hardcopy of the transcript every time they thought it occurred. They were also asked to make comments in ink in the margins noting any reasoning for their choice of codes or any uncertainties or disagreement. Space was left on the coding schedule for any new, missing, codes to be added if the checker wished. They then returned the annotated hardcopy transcript in full.

The checkers coding was found to be largely consistent with the lead researcher's. The notes on rationale made this relatively clear. Some minor code wording amendments, made to reduce ambiguity, were then applied to all transcripts retrospectively in a final coding. In one case, one code checker ascribed a different *motivation* to one driver-participant's answering throughout his transcript. This driver had commented negatively about various semi-intelligent secondary systems in his car. Whereas the lead researcher had concluded the driver found these systems 'unnatural' because of various interactional design failings, the code checker concluded this driver had a phobia of technology generally. This was a difficult issue to resolve as both interpretations were subjective. Fortunately, it was possible to contact the driver concerned, and following a telephone conversation, ascertain that it was predominantly a naturalness rather than a technological issue.

Independent checkers were also used when clustering the codes into the higher order themes in the final framework. This was done by using the A*ffinity Diagramming* technique described in Hanington and Martin (2012) originally devised by Beyer and Holtzblatt (1997). Two independent checkers, both interaction design researchers at Brunel University, were used to do this. It involved writing all the codes onto Post-It notes and grouping them individually and then discussing as a group until consensus was achieved.

In presenting the results of Thematic Analysis, it is essential that narrative meaning is preserved, and the researcher's interpretation presented transparently (Braun and Clarke, 2006). The recommended way of doing this is by using direct quotations from participants, so the themes are presented together with a selection of quotations that led to them.



	Transcripts	What it probably means	Codes	Categories	Themes
2	But doesn't one already control one's car very directly?	Does not want changes.	Satisfied with the present	Rejection of change	Restoration
3	But don't you communicate with a car very directly now?				
4	...the visual fact of not having a steering wheel may put you off.	Thinks the steering wheel is essential component of a car.			
5	I don't think I can control it more directly than I'm allowed to control it at the moment.	Thinks the current way of control is the most direct way.			
6	I actually think not much one different from the way one does it now.				
7	Removing steering? Steering with one's brain and one's eyes?	Skeptical with the radical change			
8	Well isn't that how it's going to go? There would still be loads of cars on the road? I mean 2030 is only 17 years away.				
9	...2030 that's only 18 years away...				
10	That's very sci-fi, isn't it?				
11	Or something like that – is that right?	Have not thought about a car in the future at all.	Not familiar with the technology		
12	I don't have enough confidence in that.				
13	I dunno, probably...				
14	I just don't know enough information about car development.	Not sure about the trend of car development technology.			
15	I'm not sure that we're anywhere near that at the moment.				
16	I don't know enough about it.	Not familiar with the technology			
17	I've never used Siri to create texts, for instance.				
18	It would be difficult to trust a system.	Does not trust the machine.	Rejection of technology	Rejection of technology	
19	I think you'd need a manual backup because if anything went wrong...				
20	As long as you've got manual override I don't see a problem with that.				
21	...but it could be distracting.	Too much information could be a disturbance for driving experience.			
22	it would feel OK as long as it only replied when you interrogated it! I wouldn't want it to just starting spurting information!				
23	...if you're thinking of voice control or even mind control, you're going way into the future, you just think about the future and a car does it. But how does it react to the person that walks out in front. You know you'd need to have so many sensors around that the car would be prohibitive in cost anyway.	Thinks the car with high technology is not ready to be manufactured and sold yet.			

Figure 3.4 Example of an independent code checker's alternative thematic framework

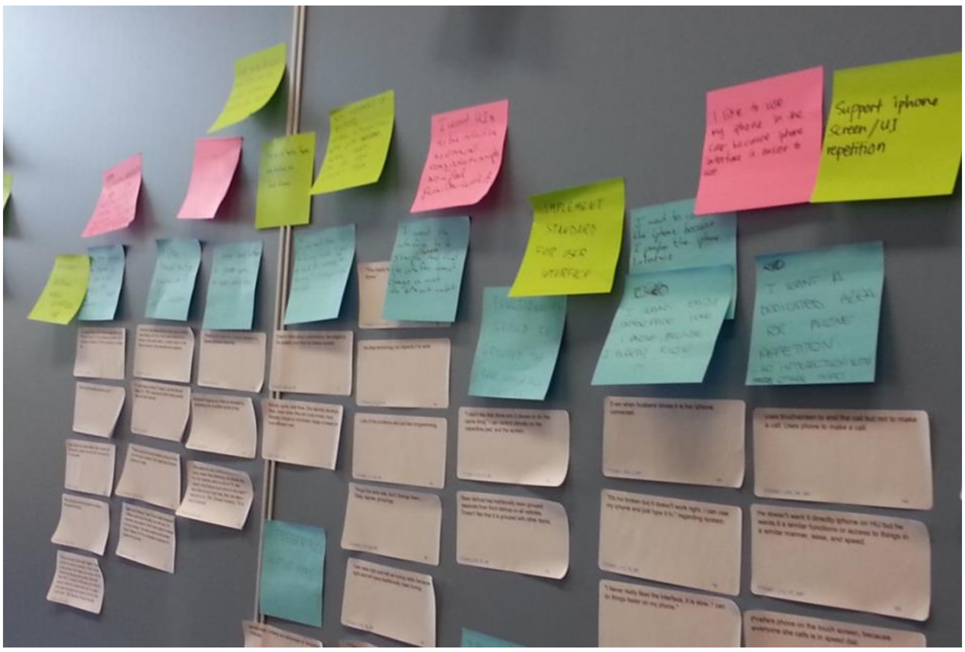


Figure 3.5 Example of affinity diagramming during independent code clustering

3.9 RESULTS

Ten men and two women aged 30–70 (mean=48; SD=14) were interviewed (mean interview time 37 minutes), followed by three further men to check no new codes arose. Regrettably it proved impossible to recruit as many females as males, perhaps because the lack of payment incentive meant car enthusiasts were over-recruited (who, stereotypically, would tend to be male) compared to the population as a whole. However, the gender balance did not have to be exactly equal for an exploratory study since the results were not intended to be generalised onto the entire driving population at this exploratory stage (Morse and Field, 1995). No new codes or themes were found in the three further 'saturation check' interviews conducted after the initial 12 had been analysed.

Ten themes are proposed which appear to characterise the construct of natural-feeling driver-car interaction. Five are **physical and control interaction characteristics** which appear to contribute to natural-feeling interaction, and five are **social and intelligent perceived behaviours** which appear to characterise natural-feeling driver-car interaction especially with more intelligent secondary car systems. These were all derived from the *Thematic Analysis.*

Quotations are presented with interviewees' gender, age bracket and car type (e.g. "F 35-40 Premium") for biographical context. Any emphasis in the transcribed quotes is the interviewee's own. Square brackets indicate minor readability edits for the benefit of the reader of this report, or to give missing context from the originating question.

Generally, drivers had quite strong *expectations* for their cars which centred on reliability (mentioned by 85% of drivers) and ability to perform one or more instrumental roles (also mentioned by 85% of drivers). Other expectations concerned 'enjoyment' and 'privacy'. However, there was no specific indication that these expectations were central to *naturalness* – it appeared more likely these were no different from the general expectations human operators have over *any* product or machine which they operate. Therefore expectation is not included as a naturalness theme.

3.9.1 THE 10 NATURALNESS THEMES

The following ten themes were found which appeared to directly concern 'natural-feeling' *characteristics* or *perceived behaviours* of cars. They divide into two main types – Themes 1 to 5 which were about *physical and control* aspects of natural-feeling secondary control interaction, and Themes 6 to 10 which were about perceived *socio-intelligent* aspects of natural-feeling secondary control interaction.

THEMES 1 TO 5: PHYSICAL AND CONTROL CHARACTERISTICS

Theme 1. Full Control and Manoeuvrability

A group of interrelated themes concerned overall control of the car. The way they were expressed suggested they were deeply implicated in natural-feeling driver-car interaction. Most drivers expected to feel in full control:

Most of the time I feel in full control of the car. And I'd be worried if I didn't. *M 40-45 Premium*

When asked to explain what full control *felt* like, the general semantics arising were those of flow, mastery, human-machine unity, and positive contextual factors:

Yes I think I feel in full control when I'm on an open road with little traffic and I'm rested, I'm not hungry, I'm not thirsty. […]… when you feel in full control you're usually feeling a degree of pleasure from the experience. I think I had an experience like that a few weeks ago […].no jams, the car was behaving perfectly. *M 65-70 Premium*

This sensation and pleasure of full control was not primarily linked to power or speed, but to contextual factors and the physical inputs into the car.

Despite the strong desire for control 'autopiloting' was frequently described, i.e. the use of such learned and familiar skills that drivers do not consciously think about their action or even recall a recent interaction. This applied to primary as well as secondary interactions.

Um well you have to use the… it's all reflex actions isn't it, it's like riding a bike or swimming *M 40-45 Premium*

Theme 2. Direct Connection

Constant mechanical (sonic and vibrational) feedback appeared to be perceived as natural.

Just by the feel of it and the noises. [...] I'm definitely one for spotting a peculiar noise sometimes and you think 'hang on a minute that's not right' *F 45-50 Premium*

Steering also acts as a 'natural' connection with the road ahead, and this metaphor appeared to extend to other secondary controls as well.

To me through the steering wheel you feel the actual roughness of the road, the camber of the road, the bumpiness of the road […] I do *like to have the sensations….* *M 40-45 Premium*

I mean I have driven cars where [the power steering feels] completely disconnected and that doesn't feel right to me, that feels like playing a computer game... M 40-45 Premium

Stop-start systems, which save fuel when cars are stationary for short periods by cutting out the idling engine, arose in several interviews as an issue of some concern. The lack of reassurance the car would restart, the 'unnatural' silence and lack of vibration appeared to undermine the naturalness of the direct link between driver and car:

I need it to keep running […] I just need to have the comfort it's just purring away and ready to go. M 40-45 Premium

I personally don't like stop-start because I would always feel 'is it gonna start?' I mean 99 or 99.9 percent of the time it probably will, but what happens if it doesn't? M 65-70 Citycar

Overall it was difficult to discern whether the 'direct connection' was to the car's mechanicals, or the road, or both, thus the title of this theme reflects a general 'reality-based' physical connection.

Theme 3. Rich Skilled Physicality

Rich physical inputs and learned-skilled input actions were described in very natural terms.

I feel driving is a craft, and I quite enjoy crafting and, um, the combined effort to get somewhere… M 40-45 Sports

I just got the knack of it of doing it really well and it just feels like you're really masterful […] that really fine balance […] it gives you a real sense of exhilaration almost. M 40-45 Premium

Quickly I'll just flick down twice… I find that very intuitive and it works an absolute treat… but I always put it back into automatic mode after that. M 65-70 Citycar

Overall this theme encompasses a sense that precise, weighted analogue, physical interactions felt more natural than digital 'clicks', lightweight-feeling interaction or binary mode selections. Inherent was also a sense that natural interactions had an instant, closely coupled 'cause-and-effect' relationship of the driver acting on the 'real world'. This led to a sense of 'craft' and in many drivers a real satisfaction in the learned mastery of multiple skills.

Theme 4. Comfort

The most mentions of any individual naturalness expectation semantic related to comfort – typically adequate comfort of seats, and maintaining appropriate of internal cabin climate, especially ventilation.

When I proceed I'm very comfortable. I do like the controls up here on my wheel. M 70-75 SUV

It feels a bit like I'm still at home, because I'm comfortable in my car, and I've only recently got out of bed. M 35-40 Luxury

Many drivers implied they expected to actively adjust the climate controls while driving. It may be that it is an expectation from older car's systems which required more active input, or that current systems do not offer adequate feedback, or do not gauge the driver's own feeling of temperature taking into account contextual factors like the weather and clothing.

I start moving straight away. I adjust the fan thing, I look for this button that demists the car, and then knock off the air-conditioned component of it. […] The fans are sometimes too loud to hear the radio which is quite annoying M 35-40 Luxury

Drivers may also *enjoy* controlling them:

The air conditioner […] and the radio I guess [would be the most important interactions]. Coz you're playing with the radio, playing with the air conditioning... M 30-35 Sports

Comfort expectations like seats, music and climate had high saliency and drivers sometimes spoke of them in natural terms. However, the link is unclear and it may be that adequate comfort is better expressed as an 'expectation' (i.e. in the centre of the framework) rather than being a contributor to naturalness directly. This is consistent with two comments that a natural-feeling car would "not be too comfortable". With the other naturalness themes 1 to 10 apparently fostering naturalness by their 'presence' alone rather than to a degree, the role of comfort has to be questioned.

Theme 5. Vehicular Usability

Good visibility and strong feedback were key expectations and strongly implicated in natural-feeling car operation, and apparently related to being in full control (Theme 1). Although vehicular ergonomic preferences were not probed in detail, this theme encompasses an apparent naturalness preference for ergonomic vehicular usability 'recommended practice' such as designing for eyes-free operation of controls, minimal distraction, and low mental workload. Other aspects of usability will need to be explored more explicitly in future studies to ascertain whether usability is a *general* expectation or one specific to naturalness.

Gauges were very salient in drivers' answers and appear to be perceived as more a natural type of feedback than 'binary' warning lights. Gauges were mentioned more than three times on average by each driver. There was a marked difference in how much, and how naturally, drivers spoke about richer analogue type of feedback (e.g. a fuel economy gauge), compared to binary type feedback (e.g. a 'check engine' warning light).

So I mean at the moment the conversation is done through your eyes. I'm looking at the fuel gauge, the clock, the speedometer, the outside temperature, the amount of fuel left but if it was sort of updating you and saying [things] to you. M 40-45 Premium.

I do keep a fairly good eye on the temperature gauge, […,] I do like to have a physical gauge to see what's going on… M 40-45 Sports

The few participants (around 30%) who had pictorial graphical displays (for hybrid energy use, systems monitoring, or built in GPS) spoke frequently and enthusiastically about their display. There were suggestions that pictorial graphical displays may be a natural way for displaying certain kinds of automotive information (along with analogue dials).

Feedback which used more than one 'mode' or 'channel' of communication, whether intended or not, tended to be more richly described – for example sound *and* vibration, or a visual indicator with accompanying sound. This is in accordance with the usual assumption that humans are a 'multimodal species' (Hancock, 1999; Wickens et al, 2013) and have a preference for multimodality in machine communication. Many drivers described, in natural terms, the constant mechanical 'chatter' from their car which in ordinary circumstances might perhaps be ignored, but becomes quite salient and informative if something is out of order:

I judge it by the vibrations and the noises and sounds, and it feels steady. […] I just familiarise myself with those [the sounds] so as long as it's consistent, it's really judged on consistency of sounds and noises, vibrations… F 45-50 Estate

I've found with all the cars that I've had, if there's something wrong, you can FEEL there's something wrong. M 40-45 Premium

I don't think I'd like a completely silent car, I'd still like a bit of noise of an engine… feeling that there is something, a power plant there. M 40-45 Sports

THEMES 6-10: SOCIAL-INTELLIGENT PERCEIVED BEHAVIOURS

Compared to the themes above, there was remarkable consensus around the five themes below, each expressed by 80% of interviewees usually several times each. These themes arose mainly from the questions and visualisations regarding *future* intelligent cars, so are more future-focused than Themes 1 to 5.

Theme 6. Acts Like a Technical Co-pilot

In this theme, many drivers described a business-like partnership, or co-pilot-type relationship (pragmatic, subordinate but respectful) when describing natural-feeling interaction with current or future cars:

I see it as a partnership […] I wouldn't get there without the car, and it wouldn't get there without me […] I don't mollycoddle it, I expect it to make progress… *M 40-45 Sports*

It would be very handy to be able to bark out instructions as if you had a co-driver *M 35-40 Luxury*

Here the theme was of a natural partnership or a subordinate but respectful relationship. The car was regarded as having quite narrow capabilities, with a technical, functional type of intelligence. If their natural-feeling future car could 'talk', drivers would expect it to talk only about its area of expertise rather than making general conversation:

It would give you more current feedback if there's something out of line, a belt's loose, radiator fluid needs topping off... *M 30-35 Sports*

There was a sense that an *intelligent* natural-feeling car could be trusted to get on with certain technical jobs on its own. For example, from an interviewee with a hybrid car:

The time it takes control of itself is the time it switches between power sources… [it feels] quite pleasing. It's taking actions for the right reasons I suppose. *M 45-50 Hybrid*

Theme 7. Humanlike Proactive Assistance

This theme achieved over five unprompted mentions per person per interview, more than twice as many as any other theme here. This theme was exemplified by the natural-feeling car *taking some action* or offering targeted assistance based on *information it had sensed and processed itself*. Examples of 'humanlike proactive assistance' include doors opening and unlocking, climate adjusting to usual settings or body temperature, and seats moving into helpful positions. The theme of driver recognition was implicit in most of these:

I presume that the door will have opened, coz it will have sensed that you're walking up to it, the door will open, the steering wheel will move out the way, the seat will move right back, you'll get in, sit down, everything will come back to your position, the seat belt will do up for you, the car will start, it will have checked all its stuff around it... M 65-70 Citycar

This theme does not include automation of the *primary* controls, which at the time of the interviews (2013) was rarely expressed in natural-feeling terms. It tended to be the repetitive, mundane, predictable secondary control tasks that were described as being 'naturally' automated in the future. Humanlike intelligence and perception appeared to be perceived as more natural. Natural-feeling assistance should mimic the actions of a competent human in that same situation, and not add the 'cognitive overhead' of having to supervise the automation:

When it's dull or when it starts to drizzle the [auto-] lights won't come on and you then have to turn them on…. To me something that's automatic should be completely *automatic, it should sense when YOU might decide to put your lights on. M 65-70 Citycar*

The lights in this car go on automatically, which I like, coz I don't have to think about it. But occasionally I get people flashing me and I wonder if it's done the wrong thing. M 35-40 Luxury

Cruise control is uncanny because it does keep the same speed even when you go down a hill and into a dip and then up the other side […] if you were just using your own brain to control the car, you would be going more slowly up the rise than […] down! M 65-70 Premium

Theme 8. Intelligent Sensing and Understanding

Most interviewees described natural-feeling cars as having improved sensing compared to current cars. This theme was defined by the sensing of environmental, mechanical or contextual parameters, and presenting that information to the driver, but *not* taking action. Typical examples of the greater sensing theme concerned mechanical parameters, tyre pressures, traffic information, road surface information, and external temperature.

Cars need a lot more sensors for what's going on. For example, you've got a thing that goes ping when it's about to freeze outside but that's just one temperature. They never tell you when it's a bit damp on the road and you might not have noticed. M 40-45 Premium

I guess [I expect] more of an intelligent computer system inside the car and more sensors to be aware of what's happening inside or outside of the car. M 30-35 Sports

Instances where cars' automation had resulted in *social* signals or actions contrary to the driver's intent, or presented irrelevant options, or misunderstood the context, often caused anger and were perceived as 'unnatural'. It appeared that 'natural' sensing therefore needs to encompass social and contextual issues, not just mechanical parameters:

The most important interaction [when waiting for a friend] is to try to bloody well unlock the door […] The car can't tell that you're picking someone up. And they [try to] open the handle and they look at you in a very aggrieved fashion! And that's irritating because the car is making you behave discourteously! M 65-70 Premium

The other thing that my friend had was automatic headlights that stayed on when you left the car. He absolutely hated those and couldn't turn them off. Every time he left the car in a car park someone would come running up to him "you've left your lights on" […] and he used to have to have an argument every time. M 40-45 Premium

Theme 9. Single Intelligent Being

Drivers tended not to perceive their interactions with their current cars as interactions with a 'whole car'. Operating a current-day car appeared to be perceived as the control of many individual 'on-board systems', rather than of control of the whole car as a single unified 'user experience'. Interaction was typically described in terms of the *physical act of control* not the resulting effect.

It's a kind of dispersed brain, across the dashboard, and all of these controls, there and there, it's like a sideways L-shape. it's not a cohesive brain like a human brain... M 35-40 Luxury

You have to use the two or three pedals to go forwards or backwards as the case may be, and I suppose every car has a steering wheel. M 40-45 Premium

This appeared to undermine the naturalness of the interaction. Only after coordination of multiple system control becomes habitual and almost unconscious, does operating a car appear to feel 'natural'. By contrast, idealised future cars were mainly described as a single system (or being) that the driver could interact with on a one-to-one basis.

As I say I think voice activation is probably the way things will go. And I think that will be it. You'll get in your car and say 'drive me to X place' […] M 45-50 Hybrid

The 'character' which drivers ascribed to their cars appeared to be derived from either the brand stereotype (e.g. "German efficiency"), or driver stereotype for that particular car (e.g. "businesslike"), or the car's physical shape (e.g. "feline") or its behavioural quirks. While such stereotypes are well known in common parlance referring to the car's speed, handling or aesthetics (e.g. Hagman, 2010; Loasby, 1995) it was an important finding that these stereotypes appear to extend to the *interface* too:

I think [my Audi is] quite a masculine car so I'd like to think it had a male voice. It's definitely a business car so it would be quite business-like. Quite grown up. M 30-35 Premium

Participants often hinted that they occasionally 'naturally' treated their car as a sentient being in some respects:

In a funny sort of way you do *want your car to feel as though you appreciate it, because in that way you hope it won't let you down. And I know that's irrational because it's only a piece of machinery. […] you almost tap the dashboard and say, 'well done, thanks'….* M 30-35 Premium

Theme 10. Vocal Information Exchange

Most participants described wanting to invoke, set or adjust a particular feature *by voice*, which was described in natural-feeling terms. It tended to be described as a short direct *command*, with the vehicle responding by action (rather than reciprocal voice). Examples were music selection, GPS destination entry, and cabin climate setting. In future natural-feeling cars however, 'Intelligent exchange by voice' was a more common theme whereby drivers described a *conversational* two-way dialogue, not resulting in instant action, but instead contributing to an overall dialogue between sentient beings or equals:

The dashboard would say 'good morning' and 'where are we going to today… and you'd say 'we're going to Holborn' […] and it would say 'certainly; sit back, relax…' M 65-70 Premium

Several drivers commented that their car already in effect had a 'voice', this being the voice of its GPS. There was no consensus about a car voice's gender, this apparently being easily influenced by their brand perception, current GPS's voice or even external technologies like Apple's 'Siri' voice. Drivers would apparently 'naturally' talk either to their dashboard or steering wheel. The overall impression was short, concise, polite, data-centred exchange.

Whereas all the other themes listed above arose largely unprompted and unprejudiced in the interviews, three of the later interview questions explicitly concerned a 'talking car' (because in the pilot interviews it had been observed to be a good way of elucidating drivers' relationship with the car). The listing of this theme at the end of the Results is therefore indicative of the possible prejudicing that led to it.

For each theme an illustrated thematic hierarchy was drawn up, illustrating some of the raw data which led to each subtheme, and the subthemes which were grouped together to form the main theme. An example is shown in Figure 3.6 and others are available on request.

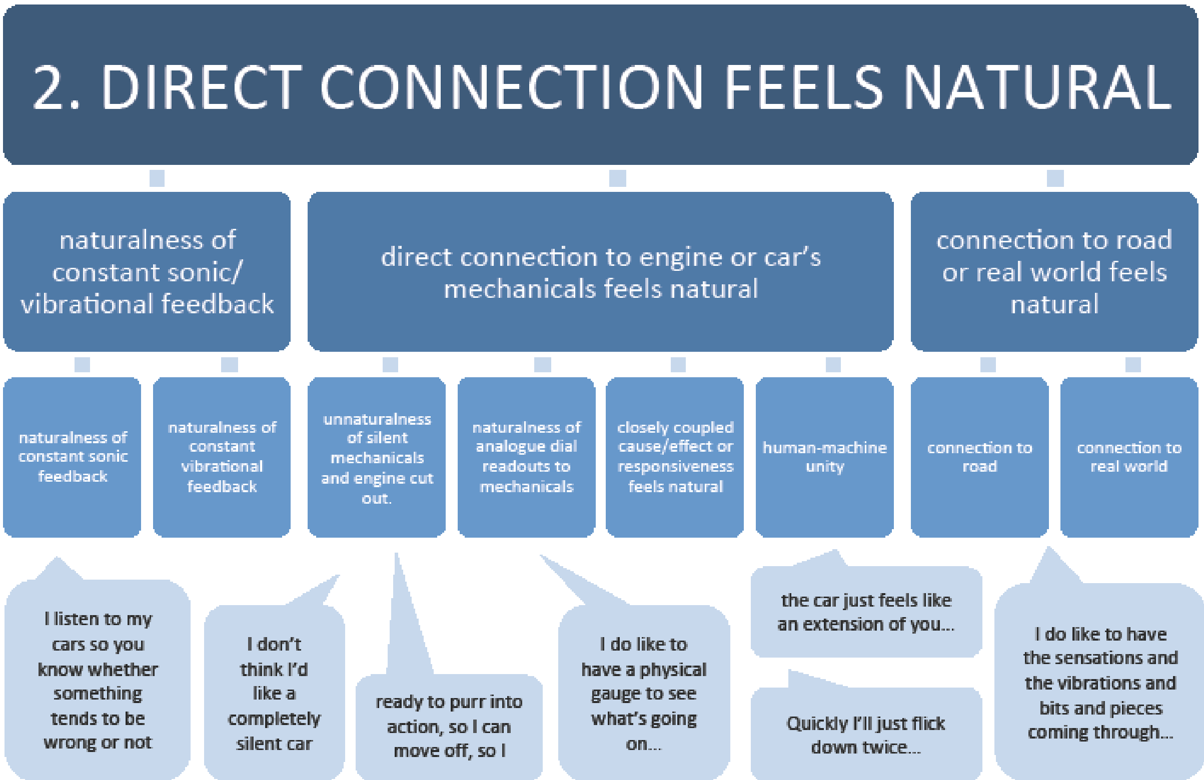


FIGURE 3.6 EXAMPLE OF THE THEMATIC MATRIXES DRAWN UP FOR EACH FINAL THEME (THIS EXAMPLE IS THEME 2 'DIRECT CONNECTION')

3.10 NATURALNESS FRAMEWORK WHICH EMERGED FROM THE CONTEXTUAL INQUIRY

The results may be summarised by the framework below (Figure 3.7). On the **right** are some interface characteristics relating to 'physical' and 'control' factors which appear to create natural-feeling interaction; on the **left** are some 'social' and 'intelligent' *perceived behaviours* of cars which appear to lead to more natural-feeling interaction. Whilst this was only a first exploratory qualitative study, there were indications in the data that the themes were bipolar in nature. This means that the presence of a theme led to perceptions of naturalness, whilst the presence of the semantic *opposite* of that theme led to perceptions of *unnaturalness*. This is discussed in detail in Chapter 7 which describes a quantitative study. At this stage it was not supposed that the themes were orthogonal (independent from each other) because so many were clearly related and distinctions were not always easy to make.

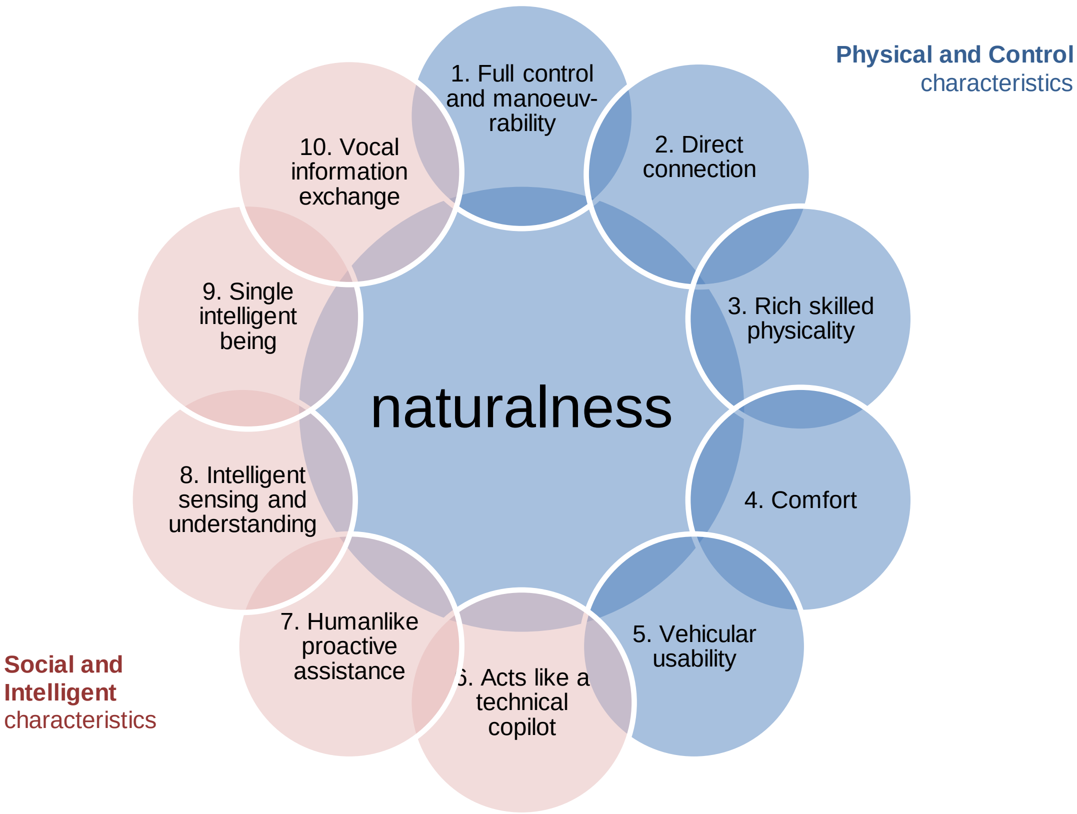


FIGURE 3.7. FRAMEWORK OF 10 CHARACTERISTICS AND PERCEPTIONS OF NATURAL-FEELING DRIVER-CAR INTERACTION FROM THE CONTEXTUAL INQUIRY

3.11 DISCUSSION

This first exploratory study's proposed framework of ten naturalness characteristics was the first step towards understanding what aspects of driver-car interaction with automotive controls feel 'natural'. The themes were independently verified by three other researchers.

Firstly, comparing the ten naturalness themes to the various **interpretations of interaction naturalness in the literature** (Section 2.6) reveals some predictable similarities but some surprising differences which may be seen as encouraging at this first stage of exploring (and justifying) *driver-car* naturalness. The most obvious similarities are between Themes 2 and 3 and the 'reality-connected' and 'skilled/whole-body' interpretations of naturalness concerning manual skills, learned expertise, and reality-based interaction (RBI). The steering wheel, throttle and ventilation fan speed knob may be viewed as classical 'tangible' interaction devices which are well mapped to their effects on the ‘real world’ and therefore more natural than an abstract keyboard or WIMP interaction style in the situations and occasions they are used. But driver-car RBI is distinctive because that reality is not imagined, distant or metaphorical – it is as real as the tarmac of the road and visceral acceleration forces acting on the body, or the sensation of cold air on the face. Perhaps the kind of common human-computer interactions studied by ‘interaction naturalness’ writers have involved little control on or in ‘the world’ and may have been intended to relate to ‘information processing’ type computing or home computing (or written at a time when cars were not considered intelligent enough to be computers). By contrast the automobile, arguably now also a computer or even robot, has a genuinely ‘real’ interaction with the road and environment – and the consequences of hindering or obscuring this reality might be dangerous or even fatal - hence why that might feel so unnatural. This finding (i.e. about grounding interaction in its real-world consequences) could also conceivably be applied to aviation, air traffic control design or design of nuclear power controllers; it also has much in common with the ecological interface design movement.

Theme 5 has parallels in the naturalness literature about ‘intuitiveness’ because it implies cognitive skill transfer (or again, learned familiarity) while the sense of ‘ease’ in Theme 1 implies both sensory-motor skill transfer (or learned familiarity) and low cognitive demand. There was a sense from the narrative (which will need to be confirmed by observations) that many natural feeling interactions happen almost unnoticed and habitually. This aspect is not often discussed in the literature (although arguably it might be implied in Wigdor and Wixon’s (2011) example of a musician playing their favourite instrument without looking at the musical score). Themes 6 to 10 provide a novel contribution to naturalness literature because – other than robotics – highly intelligent computers have not yet tended to be considered in HCI naturalness definitions. The results here suggested that naturalness may be enhanced by adopting certain human-*human* communication stereotypes and metaphors such as ‘co-pilot’, ‘helpful assistant’ and ‘speaking/listening assistant’; and by designing the car to act as a single being or entity rather than many individual computers (again a humanlike quality). This finding could also be tested in the field of robotics and AI with the aim of defining what makes robots or digital assistants interact more naturally. Overall the naturalness themes appear very context dependent and cannot be applied too generally or compared too closely to existing literature. For example, giving feedback always either ‘eyes free’ or presenting complex information ‘graphically at a glance’ rather than alphanumerically, might be well suited to an *automobile* but not desktop computing.

In terms of practical implications, many of the **Physical and Control** characteristics (i.e. themes 1 to 5) could in theory be translated fairly literally into design parameters albeit requiring further user-generated interpretation, for example, the exact weight and tightness of controls in ‘Rich skilled physicality’ which might be obtained through a Kansei study (Jindo and Hirasago, 1997). As an example, it would appear that a driver interface characterised by rotary controls rather than touchscreens, a tight and weighted precise feel as opposed to digital clicks, multimodal rather than unimodal feedback, and needle gauges instead of warning lights, would probably feel more ‘natural’ to use. Refinements and electronics in modern cars have perhaps created a less visceral, less ‘natural’ overall interaction with fewer mechanical connections. Interviewees’ perception of current steering feel as highly ‘natural’ is contradictory because steering feel (now almost always power assisted) is highly mediated by electronics and actuators (Autocar, 2013). Perhaps there is potential to ‘synthesise’ naturalness physically, or perhaps naturalness is a fluid and unstable concept, changing over time or with familiarity. Drivers’ repeated narratives about ‘autopiloting’ (i.e. semi-conscious operation of controls) suggest automaticity, often viewed as a contributory factor in accidents (Ranney, 1994) is in some ways ‘natural’. Secondary controls designed to be operated ‘unconsciously’ may therefore feel natural to drivers. Further studies are required to ascertain if comfort is primarily a modern ‘expectation’ rather than a naturalness contributor. There was insufficient evidence from this study to judge that.

Some of the **Social and Intelligent** characteristics of the intelligent car (themes 6 to 10) could in principle be engineered today on a superficial level (and arguably already are in some cases); for example assistive seats which allow easy entering and exiting, and copilot-like delegation of routine technical tasks to an electronic assistant. However, their ultimate manifestation may have to wait until technology is capable of demonstrating ‘intelligence’ rather more robustly – for example a voice system with full natural language understanding and conversational ability, and socially intelligent ‘sentient’ awareness of driver context, mood and intent. As argued by Norman (1990) such ‘socio-intelligent’ characteristics can perhaps only be executed successfully when better technology becomes available. The data from this study suggested an ‘intelligent’ feature performing imperfectly or in a ‘machinelike’ way appeared to be quickly perceived as ‘stupid’ or ‘impolite’ (as in Cooper et al, 2012).

There may be challenges to implementing all ten naturalness characteristics concurrently because in some respects they contradict. For example, despite road noise and mechanical feedback being steadily reduced in recent years for ‘comfort’ reasons (Walker et al, 2006), the data suggested such sonic and vibrational ‘direct connection’ feels natural. ‘Visibility’ is a ‘natural’ expectation but so too is ‘privacy’ which often implies ‘invisibility’. ‘Rich skilled physicality' is not necessarily compatible with 'vocal information exchange'. Some future fully self-driving arrangements would seem contrary to the naturalness themes which suggest humanlike assistance. Automation of the primary controls was, in 2013 at least, rarely expressed in natural-feeling terms. This perception may perhaps change in future years once drivers become habituated to it – as with power assisted steering. Another contradiction concerns Theme 10 'vocal information exchange'. This appears to contradict reports (and observations in other studies in this research) that long-term use of car speech control is very low (Weinberg, 2012). Perhaps this is because many 'usability' characteristics (Theme 5) are currently lacking in 'push to talk' or command-word systems; perhaps current systems also lack the 'sentient' ability necessary to build shared understanding and cannot recover from errors in humanlike ways. Speech may also be insufficiently integrated for the driver to perceive they are talking to a single 'whole car' (Theme 9).

3.12 Limitations

The limitations of narrative methods are well known; however a narrative study was felt necessary to achieve triangulation (see Chapter 6). While demographics were purposely balanced in terms of the cars the selected interviewees owned and drove, and for driver age bracket, they were not balanced for gender (with a 10:2 bias towards males, probably as a result of the lack of payment incentive attracting car enthusiasts). The study design was intended to be naturalistic hence drivers being interviewed inside their own cars, but their consequent over-familiarity with those cars may have prejudiced answers towards deeply entrenched perceptions or feelings, and prevented less familiar/more novel perceptions being elicited as for example when someone tries a new car for the first time, or rents an unfamiliar car on holiday. Although saturation appeared to have been achieved relatively quickly with the 12-15 interviewees sampled, more interviewees might have permitted this judgment to be taken with more confidence.

3.13 Conclusion

Two pilot studies were initially conducted in order to develop a suitable semi-structured questionnaire schedule using the 'domain specific language' of drivers. Fifteen drivers were then interviewed in depth about their automotive interactions and relationships, with emphasis on naturalness issues, using ethnographic and Contextual Inquiry interview techniques which were judged to achieve data saturation in Thematic Analysis. Interpretations were kept transparent and checked by external code checkers. Ten characteristics of natural-feeling driver-car interaction were proposed and arranged into a framework of ‘two halves’. At this exploratory stage they include reference to primary controls as well as secondary controls to give overall context. Some potential challenges and opportunities have been suggested.

Narrative methods such as interview have limitations and some have questioned their validity in design research (e.g. Dunne, 2012). Since interviews took place in participants' own cars, the interaction scenarios used in the interviews may also have been rather biased towards the interaction style and stereotypes of that car. Perhaps overfamiliarity and automaticity of control use will have affected naturalness perceptions. Subsequent studies should therefore seek alternative elicitation methods to the narrative/interview used here, and using cars and car controls that are less familiar to the participants, in order to maximise data capture at this exploratory stage and provide a valid grounding for any naturalness framework.

CHAPTER 4

AN EXPLORATORY DESIGN WORKSHOP WITH CAR DRIVERS

4.1 AIM

The previous study elicited a narrative interview-derived framework for natural-feeling interaction between ordinary drivers and their cars' controls. Interview-based narrative methods have some predictable limitations and biases. Chiefly, what people *say* they might do or feel about interactions, is not necessarily what they *actually* do or feel when immersed in that activity in real life situations (Banks et al, 2014b; Ylirisku and Buur, 2007; Gross, 2012). Some evidence suggests that such activity needs to be *replicated* for the purposes of *participatory study* or *observation* (Dunne, 2012; Beyer and Holtzblatt, 1997). In several studies of 'natural' human-computer interaction, Reeves and Nass (1996) state *"we should be suspicious of verbal responses… many of the most important reactions and responses of users are not conscious, and hence not available for verbalisation"* (p35). Similarly, in a study about automotive cabin noise Walker et al (2006) caution that *"the level of insight needed to drive design decisions cannot always be provided… by simply asking consumers"* (p177). Using a second, different method to the first study would also permit triangulation (Morse, 1991) if a third method could be found for the subsequent study. A method was therefore sought to develop a second, independent inductive ('bottom-up') framework of natural-feeling driver-car interaction, preferably replicating some practical activity related to secondary system control and capturing 'instantaneous' perceptions rather than post-reflected narrative. The research question was broadly the same as the previous study, but now needed to be restricted to secondary controls only:

What are the component characteristics and dimensions of natural-feeling interaction between ordinary drivers and automobile secondary controls and in what circumstances does it tend to occur?

4.2 RESEARCH METHODS

Numerous design researchers (e.g. Ylirisku and Buur, 2007; Dunne, 2012) have suggested that participatory *artefact-centred* and *visualisation* activity is a suitable complement to narrative methods. Human-centred design research methodology (e.g. the 100 methods in Hanington and Martin, 2012) suggests several practical exploratory 'workshop' methods that can potentially elicit 'instantaneous' perceptions through creative activity. In theory, these are less prone to narrative bias (Patton, 1990) and according to recent neuroscientific research may elicit more ‘holistic contextual’ perceptions compared to the ‘abstracted self-referencing’ schematic type knowledge which may come from narrative (McGilchrist, 2009). Conducting at least part of the study in a *group* setting might give more opportunities for discussion, negotiation and insight than would be possible with just one participant-driver.

A method selection shortlist was compiled by scoring all the human centred design methods in Hanington and Martin (2012) and Ylirisku and Buur (2007) against the following seven criteria, which were selected by the researchers in accordance with the research objectives:

1. Must be qualitative and exploratory in nature (being early ‘first approach’ research);
2. Must involve ‘hands-on’ artefact-focused activity relevant to car secondary controls;
3. Must have potential to elucidate ‘private, silent’ interactions;
4. Should capture immediate perceptions *during* interactions, not post-reflected narrative;
5. Suitable for design research ‘discovery’ and ‘problem definition’ stages;
6. Could be readily applied at the physical scale of the automobile dashboard;
7. One method should involve some group (social/negotiated) activity.

Participant observation, questionnaire and interview based ‘self-reporting’ methods were excluded because they had been selected for the previous and subsequent studies. This resulted in a shortlist of six methods which are outlined below:

4.2.1 ‘Think Aloud’ Protocol

‘Think Aloud’ Protocol (TAP) testing was first described by Ericsson and Simon (1984) as a way of eliciting users’ thoughts, mental models, reflections and affective responses during interactions (e.g. Goodman et al, 2013). Users are literally asked to ‘think aloud’ during or just after interactions of interest (Makri et al, 2011). TAP is often employed in late stages of information architecture and website design but is rarely seen in published automotive literature, an exception being Banks et al (2014b). It has on occasions been used to seek to understand ‘natural behaviours’ (Makri et al, 2011). There are typically three variants of TAP: Concurrent, Retrospective and Cooperative. Concurrent TAP, most used in interface design, aims to understand interactions *as they occur*. TAP can be time intensive and is typically undertaken on just 5-12 users while being audio or video recorded (Makri et al, 2011). It is typically analysed by thematic analysis (e.g. Braun and Clarke, 2006). Ericsson and Simon’s (1984) original TAP involved a simple single prior instruction to ‘think aloud’ with no ‘probing’ allowed. Critics of this original method suggest it can be rather silent and unnatural (Olmsted-Hawala et al, 2010) or that it only reveals *actions* not the underlying thoughts behind them (Makri et al, 2011). Perhaps because of this many design researchers have adapted the method to 'probe' relevant interactions, thoughts and body language as they occur. TAP with 'concurrent probing' is usually considered valid providing the 'probing' does not go beyond the 'task in hand' (Goodman et al, 2013).

4.2.2 EXPLORATORY DESIGN WORKSHOPS

Design workshops may be thought of as the facilitation of a group of product users to undertake various hands-on, engaging, reflective activities in order to gain deeper understanding of customer needs, meanings and perceptions, for the purpose of product or service design. Some design researchers have suggested that the type of human understanding essential for successful product design is accessed *only* through such 'creative play' (e.g. Ylirisku and Buur, 2007). Thoughts and feelings arise as a direct result of stimulating exercises (often with physical artefacts) and may be elicited instantaneously. Typically, *exploratory* design workshops are used for generative design or evaluation to create an empathic knowledge base about a relatively unknown market area (Hanington and Martin, 2012) and for exploratory design research (ibid). A common way of conducting a workshop is to ask participants to create a representation of an 'ideal product' from various physical components. Involving users as 'co-creators' in this way, can produce ideas that are considered creative and highly valued (Kristensson, 2008). Exploratory design workshops have occasionally been used in automotive product design (e.g. Cycil et al, 2014).

4.2.3 'BREACHING'

Breaching, first described by Garfinkel (1967) in social psychology, seeks to understand people's reactions to *violations* of social or design norms (e.g. Baharin et al 2013). The theory underlying breaching is that people are not always consciously aware of the 'unwritten rules' that build up around interactions, which may only accessed by 'violating' them. A 'breaching exercise' can be a useful extension to an exploratory design workshop (Degen, 2014) in which participants are asked to create a 'worst possible' product or service.

4.2.4 FLEXIBLE MODELLING

Flexible modelling is a 'hands-on' participatory method often used within exploratory design workshops for generative, exploratory or evaluative research by helping users express their needs and desires in physical form. Given a kit of components, an engaging practical task, and focused facilitation, valuable insight can be provided into interface configurations (Hanington and Martin 2012). The associated discussions provide an opportunity to enquire as to the reasoning behind users' design decisions and perceptions. Kits may comprise general 'non-specific' components for more open-ended design tasks, or 'familiar specific' components when the components are relatively fixed but not their arrangement (ibid).

4.2.5 *FOCUS GROUPS*

Focus groups are a well-established form of 'collective interview' used in market research, and increasingly in applied psychology (Coolican, 2009) comprising professionally facilitated discussions on a topic of interest usually with a small group of non-experts. The purpose is usually in-depth exploration of a topic about which relatively little is known (Stewart and Shamdasani, 2014). Focus groups are sometimes used in design research with an artefact stimulus, for example when a consumer product needs improvement or redesign (Hanington and Martin 2012). The key difference from interviews is the *group* stimulation and social 'negotiation' of responses to questions, with relatively flexible questioning, which may elicit insights which might not otherwise emerge (Coolican, 2009). Focus groups are usually highly focused on a single research question to 'exhaustively' explore relevant thoughts and feelings (Stewart and Shamdasani 2014). Like ethnographic interviewing, focus group discussions commonly begin with open-ended 'grand tour' questions (Spradley, 1979) that seek to obtain participants' overall orientation towards the topic. A pre-exercise is often given to participants to be completed before they attend the research, as a 'sensitising' exercise to the topic (Stewart and Shamdasani, 2014). Full transcription captures people's whole responses which can then be interpreted rigorously using thematic or conversational analysis (ibid). Focus groups are probably used in industrial automobile interface design but their results are rarely published, probably for commercial reasons (Schmidt et al, 2010).

4.2.6 *FUTURE FICTION*

Future fiction, or *science fiction prototyping*, is the creation and use of science fiction story, film or comic, typically based on real science and feasible technology, to explore the implications and potential usage of future technologies (Johnson, 2011). Ylirisku and Buur (2007) and Dunne (1999) strongly suggest creating 'realistic but compelling' scenarios to display a future that people may have difficulty imagining otherwise. In the automotive domain, Gaspar et al (2014) suggested using 'imagined scenarios' to make laboratory 'bench testing' of controls more realistic, while Gkouskos et al (2014) created two 'pre-designed futures' to gauge driver needs relating to future vehicles.

One interview question in the previous study (also described in Ramm et al 2014) had asked drivers to visualise what they imagined happening when getting into their cars to start a future journey in 2030. Many drivers described a type of 'natural vocal exchange' with their automobile. Their descriptions of what their automobile was likely to *say*, and what they would say to it in response, had appeared to provide insight into 'holistic' perceptions of future automobiles, for example what personality, status and agency they attributed to them.

4.3 KEY METHODOLOGICAL DECISIONS

Each of the five shortlisted methods was considered to meet at least five of the seven criteria above in its standard deployment. When combined with the key operational decisions and adaptions below, each was considered to meet six of the seven criteria. Each method also represented an opportunity to ask the research question in a subtly different way. Such variety of methods and stimulus change can be more stimulating for participants (Cornish et al 2009). Therefore, it was decided the workshop should be a mix of **all six methods** above with the methodological decisions and adaptations described below.

1. '**Think Aloud' Protocol** is normally used in prototyping and late stage testing, but the literature did not suggest it could not be used for exploratory means. 'Concurrent Probing Think Aloud' appeared to be one of the few methods which could potentially elucidate complex 'private silent' interactions with automobile controls. Therefore it was selected.

2. An **exploratory design workshop** was considered to be a logical choice because of the need to access immediate and minimally post-rationalised feelings and perceptions about machine interactions and interface arrangements, and the desire for some 'social' activity.

3. Since most of an automobile's secondary interaction takes place at its dashboard, with mainly familiar controls, a **flexible modelling** exercise with both 'specific' and 'non-specific' components appeared logical. A session was envisaged whereby a group of drivers would create a 'natural-feeling dashboard' on a table-top template using automotive controls extracted from a variety of real cars. The resulting creations could be photographed and participants asked to discuss their choice of components and layouts. A **breaching exercise** would be similar but the objective would be to create the most 'unnatural-feeling' dashboard.

4. **A focus group** format was considered the most appropriate way of achieving further 'deep enquiry' about the topic of interest. It would allow both stereotypical and unexpected responses to be thoroughly explored. The focus group stood out against other academic methods of inquiry because it permits inclusion of *artefacts* (Stewart and Shamdasani 2014).

5. Although not 'Contextual Inquiry' *per se*, this study was informed by the principle of **contextual faithfulness**. It was decided that one focus group, one 'Think Aloud' activity, and the future fiction scenario should all take place *inside* a real (parked) automobile.

6**.** This study was intended to be relevant to the future, so some perceptions of un/natural-feeling secondary systems in *future* and *'intelligent'* automobiles were sought. A brief 'talking car' **future fiction** would be devised. Participants, seated in a car, would be asked to imagine it was 'talking' to them. This scenario was chosen because in the previous study many drivers had imagined natural language as a likely way of controlling cars in the future.

4.4 Objective

The study objective was to use the six activity-based exploratory and visualisation methods described above with small groups of automobile drivers in a laboratory-based workshop environment to elicit interaction perceptions (Gross, 2012) relevant to the research question. Ecological validity would be maintained by using real cars where appropriate and real car controls in all the other exercises. Perceptions would be captured instantaneously by using Think Aloud and focus group techniques, flip charts, photography and audio recordings.

4.5 Study design

4.5.1 Trial workshop

The methods described above were tested and refined in a trial workshop involving three automobile-owning product designers who were not directly involved in the research. Involving designers in early stage workshops can be an efficient way of optimising their design (Cycil, 2016). Small groups of around four participants had been envisaged, in order to create the necessary intensive activity and discussions. The trial however suggested that having more than two participants per workshop would make its many activities extremely difficult to manage and record, especially those conducted inside the automobile and those involving negotiation and discussion. It also suggested that groups of three or more people may inhibit the honest sharing of thoughts and perceptions. Furthermore, the six activities could not be completed in less than four hours with the three participants in the trial workshop, causing fatigue. Accordingly, it was decided that the workshops would have just two participants each. The trial workshop experience suggested that two participants, using all six methods facilitated efficiently, would take about three hours. This duration was considered appropriate in terms of human resources, efficiency and ethics.

4.5.2 Workshop session plan

Using the knowledge and experience gained from the pilot study the workshop was designed and run according to Table 4.1. Table 4.2 gives some examples of the chosen ‘prompt questions’ and ‘probing questions’ devised and refined during the trial workshop. Figure 4.1 shows one of the ‘hands-on’ sessions taking place. The messages for the speaking car exercise were typed into an iPad speech synthesiser. These six messages covered a full range of ‘social’ to ‘technical’ subject matter (for example asking about comfort breaks versus reporting minor oil level drop in the engine) and ‘humanlike’ to ‘machinelike’ in delivery (determined by varying the speech style and language from humanlike/natural to robotic/machinelike). The topics of these scenarios were derived from the previous chapter’s study, one part of which had asked drivers to imagine what a car journey would be like in the year 2030. Most drivers had spontaneously provided examples of a futuristic ‘speaking car’.

Session (duration)	**Methods Used (location)**	**Activity**	**Props/Artefacts provided**	**Data Collection Mechanism**
Introductory talk, housekeeping, waiver signature, guidelines about respect, openness, honesty				
1. Sensitisation (20 minutes)	Focus Group (table)	Discussion about sensory memories of first driving experiences (previously emailed)	None	Flipchart paper, audio recorder with verbatim transcription
2. Operating loose controls (15 minutes)	Think Aloud (standing)	Participants asked to operate various loose automobile controls, while being asked how they feel and what feels un/natural.	Two tables containing large stock of car controls	Audio recorder with verbatim transcription
3. Natural dashboard creation (25 minutes)	Exploratory Design Workshop with Flexible Modelling (table)	Imagining, discussing and creating the most ‘natural-feeling dashboard’ from loose controls on a table top template.	Two tables containing large stock of car/non-car controls AND a bench with a dashboard template drawn on it	Audio recorder with verbatim transcription; camera to photograph co-creations
20 MINUTE REFRESHMENT BREAK				
4. Unnatural dashboard creation (Breaching) (15 minutes)	Exploratory Design Workshop with Flexible Modelling (table)	Breaching exercise. Imagining, discussing and creating the most ‘unnatural-feeling dashboard’ on the same table top template.	Two tables containing large stock of car/non-car controls AND a bench with a dashboard template drawn on it	Audio recorder with verbatim transcription; camera to photograph co-creations
5. Operating controls in a real automobile (25 minutes)	Think Aloud with Focus Group style discussion (in-automobile)	Operating various fixed and loose controls in a real car. Asking about Un/natural feelings and effects of context on perceptions.	Real car parked inside the laboratory with secondary controls powered up	Audio recorder with verbatim transcription; camera
10 MINUTE BREAK				
6. Intelligent automobile (‘speaking automobile’) (25 minutes)	Future Fiction with Think Aloud style probing (in-automobile)	Audio-based future fiction. Automobile appears to be voicing various messages. Participants asked how it felt and aspects of un/naturalness	Real car parked inside the laboratory; iPad speech synthesiser	Audio recorder with verbatim transcription
10 MINUTE FEEDBACK AND PLENARY SESSION, VALIDITY TESTING EXERCISE, END.				

TABLE 4.1: SESSION PLAN FOR THE WORKSHOPS SHOWING THE VARIOUS ACTIVITIES

Session	Main prompt questions used (ie consistently for every group)
1	*What were the sensations of first using an automobile's controls? What is the general sensory experience of car controls these days?*
2	*How do these controls feel, look and sound? Which are most suitable for an automobile?*
3	*What does each element represent? What feels natural about it? Was there anything you would have liked to include? Please explain your choices of components, layouts and materials.*
4	*What does each element represent? What feels* unnatural *about it and why? How would you describe the differences to the* natural *dashboard from previously?*
5	*How does it feel to use? What do you imagine the automobile is doing in response? What feels natural or unnatural about it? How would being in a moving automobile affect the feelings of unnaturalness and naturalness?*
6	*How did it feel to hear that message? Did it feel natural or unnatural? What would be your reply? What personality should it have? How could an intelligent future automobile still behave naturally?*

TABLE 4.2 PROMPT QUESTIONS USED DURING THE SIX SESSIONS

FIGURE 4.1: OPERATING VARIOUS CONTROLS BY HAND DURING A 'THINK ALOUD' SESSION

4.5.3 VENUE

The venue selected was a large windowless research laboratory which has a complete test car permanently parked inside it (a 2002 Jaguar S-type). This was to keep the workshop's exercises psychologically grounded in the subject of automobiles and driving while minimizing effects of scenery, weather or temperature and external distractions.

4.5.4 SOURCING OF THE STOCK OF FLEXIBLE MODELLING COMPONENTS

A stock of secondary and infotainment automobile controls was required for the flexible modelling element of the exploratory design workshop, and the initial Think Aloud session. The choice was made by listing all the common input/output devices used by automobiles' secondary systems (from Kern and Schmidt, 2009, with amendments described in Sections 1.3.5 and 1.3.6; see Figures 1.1 and 1.2) and sourcing three examples of each from several automobile companies and an online auction site. No automobile brand names were visible.

FIGURE 4.2: TABLE CONTAINING SOME OF THE AUTOMOBILE AND NON-AUTOMOBILE COMPONENTS

So as not to restrict participants' choice to automobile stereotypes at this exploratory stage, 10 *non*-automobile controls were also provided – including a domestic light switch, computer game controls, a room thermostat, and calculators. A 'materials samples' collection was also made available comprising 40 samples of materials, exhibiting a wide variety of colours and textures. Blank Post-it notes were provided to allow participants to reference unavailable components, materials or intangible concepts. A year's copies of the weekly 'Autocar' magazine was made available to enable visual automotive imagery to be added if participants wished. Part of the selection is shown in Figure 4.2.

4.5.5 SCHEDULING

A total of six workshops were scheduled. This number was determined by iterative saturation checks. After each workshop a cumulative calculation was to be made showing the total number of new codes added, in order to judge if saturation had been achieved.

4.5.6 REFLEXIVITY

In qualitative research, it is recommended to reflect on experimental conditions and researcher intrusion to consider what biases may arise (Coolican, 2009). Humans also have a natural tendency towards certain common perceptual and cognitive biases (Dvorsky, 2013). A psychologist from Oxford Brookes university helped identify potential biases likely to occur in the proposed experimental conditions (considering the categories of the 'physical', 'social' and 'technological' environments – a common form of analysis used in business research). Measures were agreed relating to venue set up, automobile use, activity ordering and focus group topic guides, to minimise and mitigate against potential biases.

4.6 SAMPLING AND RECRUITMENT

There are few rules for sample size in qualitative exploratory inquiry. A small-scale in-depth study of this (very specific) phenomenon was considered to be a suitable strategy (Patton, 1990). Ten to twelve participants were sought for similar reasons to the previous study. Saturation was judged to be achieved when no more unique data 'codes' were obtained from any subsequent participant (Mason, 2010). 'Purposive sampling' was chosen, a common method in focus group recruitment (Coolican, 2009). This aimed to bring together a selection of people who are representative of the specific subgroup of interest. A 'call for participation' was placed on the Brunel University staff and student intranet in July 2014 asking for drivers aged 25 to 75, i.e. excluding very young and very old drivers who may have perceptual shortcomings (McGwin and Brown, 1999). Possession of a car and driving license were essential criteria. Physical characteristics were not controlled for because it was a small exploratory qualitative study aiming simply for verbal code saturation. A £20 voucher was offered as incentive, because of the three-hour time demand.

4.7 METHOD

The session was run according to the session plan in Table 4.1 using the prompt questions listed in Table 4.2. Refreshment breaks were scheduled. A professional audio recording device was used to capture all voices for full transcription. The lead researcher kept the timing to schedule and facilitated all the sessions. An assistant took photographs. Participants were told simply that the study was about automobile controls, but not provided with information about the 'naturalness' focus until afterwards, to avoid prejudicing. An example of the raw data is shown in Figure 4.3. Full transcripts are available on request. At the end, a small face validity exercise was conducted to test the themes of the previous study, described more fully in Chapter 8.

4.8 DATA ANALYSIS

4.8.1 SATURATION DETERMINATION

The first workshop generated 99 codes. Performing the second workshop added 47 more unique codes. The third workshop added 29 more. The fourth workshop added 4 more, making a cumulative total of 179. The fifth workshop did not generate any new unique codes. Accordingly, it was decided saturation had been achieved at 179 codes and no further workshops needed to be conducted.

4.8.2 CODING ANALYSIS

Thematic Analysis (Braun and Clarke, 2006) was used for all the workshop transcriptions because it had been successfully applied to qualitative automotive data before (e.g. Burgess et al, 2013; Ramm et al 2014) and allows rich interpretation and pattern identification from perceptual data (Saldaña, 2015). In addition, because so much data were obtained (13 hours in total, equating to over 200 pages of transcript) some basic **Content Analysis** (Krippendorff, 2004) was conducted, counting occurrences of codes to roughly estimate their strength (see Figure 4.4 as an example). A phenomenological approach to the analysis was considered more appropriate than a social constructivist approach because this research was more about the subjective idiosyncratic perceptions of individuals (Stewart and Shamdasani, 2014) rather than the social negotiation of those thoughts or perceptions. The 'bottom-up' analysis used principles of 'grounded theory' (Glaser and Strauss, 2009).

Following the procedures recommended by Braun and Clarke (2006) the full audio recording was listened to just after each workshop and tentative pre-coding reflective notes made on the full paper transcriptions in the 'eclectic' manner (Saldaña, 2015) coding fully not partially (i.e. 'splitting' not 'lumping'; ibid). An example of the raw data is shown in Figure 4.3. Data were then coded at a concept level (usually 'X is associated with natural-feeling interaction') in the 'pattern coding' manner (Saldaña, 2015). Pattern coding allows meaningful and parsimonious explanations of large amounts of verbal data (ibid). Coding was done on paper because for small in-depth studies it can give more control and preserve nuances (Saldaña, 2015) but codes were summarised in a large spreadsheet, an image of which may be seen in Figure 4.4. All statements apparently relating to naturalness or unnaturalness were coded. Their frequency of occurrence was counted across all sessions to gauge strength and exclude rare codes. Other statements were ignored.

And I'll just ask you to, if you'd take a seat again and if you can just talk me through roughly what you've got on your dashboard. …. So perhaps if we start with you on your side, so if you can explain why that's there and what it is about it that appeals to you maybe. What made it feel natural.

F: Well the seat controls I guess the only person who should be controlling, certainly the driver's seat is the driver. So it makes sense to have that to the, you know, only accessible to the driver on the right-hand side.

I: Yes.

F: Sort of the same with the window switches, they're easily locatable that's where, you know, you normally have them on like the arm don't you of the door.

M: And the other thing about the seat switches is, is putting them on the side of the seat by the door means they're really to adjust when the car's stationary. But they're difficult to knock accidentally while you're driving.

I: Okay.

M: So then there's a safety aspect to that location as well.

I: Yes so where would you have them?

M: On the side of the seat so you can get to them by opening the door, when you shut the door, oh sorry, just hit the dog, when you shut the door they're automatically locked away.

I: Right and the steering wheel, I mean that probably speaks for itself but anything you want to say about it, a feel of it or the look of it in terms of its natural feeling-ness?

F: I don't know really, steering wheels, it's got a basic function really hasn't it, I suppose it should feel quite nice to hold, you're saying like the one in your first car was quite sort of thin and quite metallic, so it might be quite, it's quite nice to have something chunky.

I: So yes give me some adjectives. Chunky.

F: Robust.

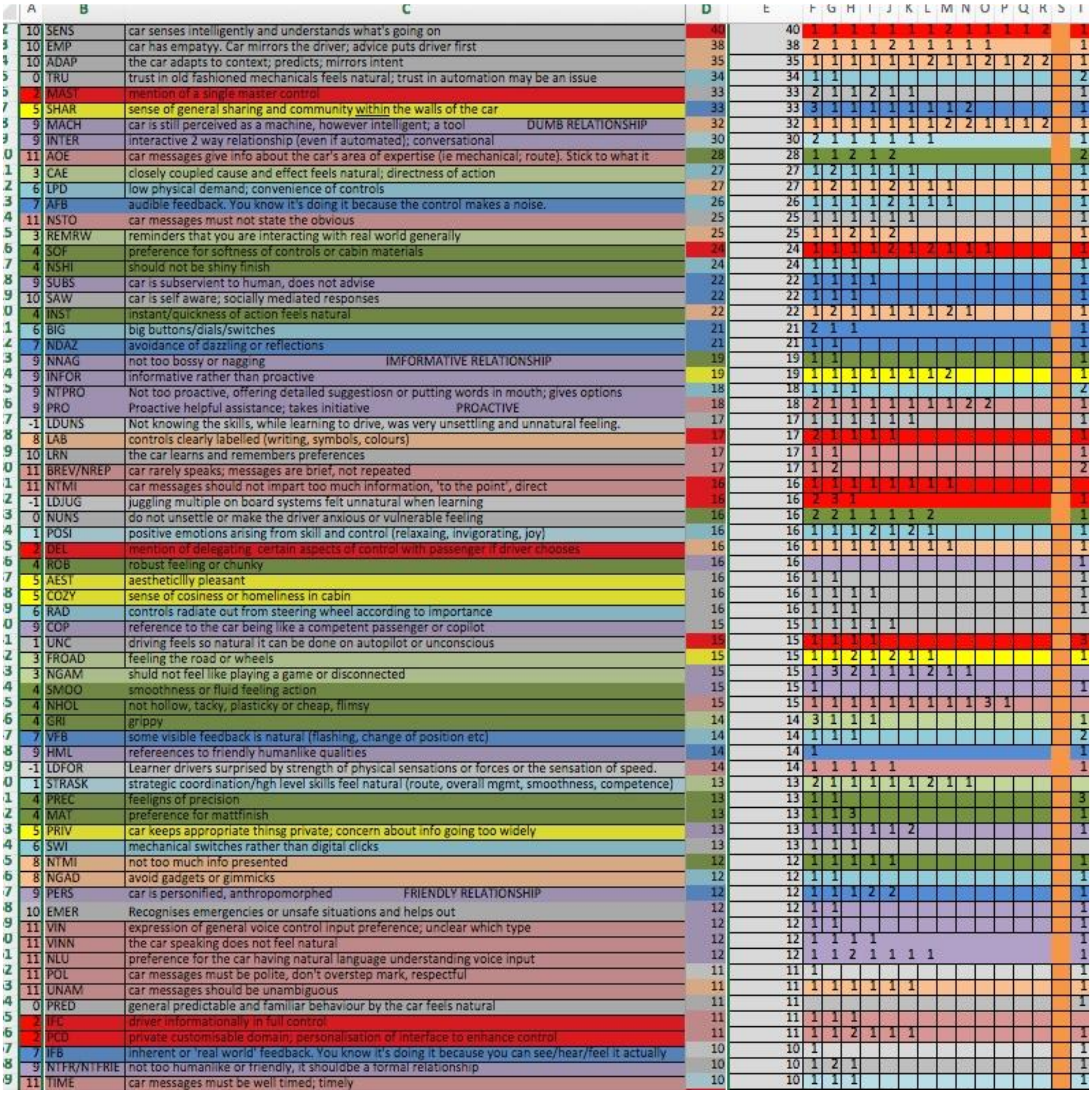
M: Well it's got to fit nicely in the hands and have a relatively high friction surface so it's easy to grip you don't have to hold it really hard.

I: Yes. Yes.

FIGURE 4.3 EXAMPLE OF THE RAW DATA (IN THIS CASE A DISCUSSION OF THE 'NATURAL DASHBOARD' ARTEFACT MODELLING EXERCISE).

4.8.3 THEMATIC ANALYSIS CLUSTERING

The 179 codes were grouped into logically similar clusters (that is sharing the same semantic meaning in the opinion of the lead researcher and one other independent clinical psychologist researcher from Oxford Brookes university) in an Excel spreadsheet. A screenshot of part of the spreadsheet is shown in Figure 4.4 from the early stages of tentative clustering where colours have been used to link codes with similar meanings.



A	B	C	D
10	SENS	car senses intelligently and understands what's going on	40
10	EMP	car has empatyy. Car mirrors the driver; advice puts driver first	38
10	ADAP	the car adapts to context; predicts; mirrors intent	35
0	TRU	trust in old fashioned mechanicals feels natural; trust in automation may be an issue	34
2	MAST	mention of a single master control	33
5	SHAR	sense of general sharing and community within the walls of the car	33
9	MACH	car is still perceived as a machine, however intelligent; a tool DUMB RELATIONSHIP	32
9	INTER	interactive 2 way relationship (even if automated); conversational	30
11	AOE	car messages give info about the car's area of expertise (ie mechanical; route). Stick to what it	28
3	CAE	closely coupled cause and effect feels natural; directness of action	27
6	LPD	low physical demand; convenience of controls	27
7	AFB	audible feedback. You know it's doing it because the control makes a noise.	26
11	NSTO	car messages must not state the obvious	25
3	REMRW	reminders that you are interacting with real world generally	25
4	SOF	preference for softness of controls or cabin materials	24
4	NSHI	should not be shiny finish	24
9	SUBS	car is subservient to human, does not advise	22
10	SAW	car is self aware; socially mediated responses	22
4	INST	instant/quickness of action feels natural	22
6	BIG	big buttons/dials/switches	21
7	NDAZ	avoidance of dazzling or reflections	21
9	NNAG	not too bossy or nagging IMFORMATIVE RELATIONSHIP	19
9	INFOR	informative rather than proactive	19
9	NTPRO	Not too proactive, offering detailed suggestiosn or putting words in mouth; gives options	18
9	PRO	Proactive helpful assistance; takes initiative PROACTIVE	18
-1	LDUNS	Not knowing the skills, while learning to drive, was very unsettling and unnatural feeling.	17
8	LAB	controls clearly labelled (writing, symbols, colours)	17
10	LRN	the car learns and remembers preferences	17
11	BREV/NREP	car rarely speaks; messages are brief, not repeated	17
11	NTMI	car messages should not impart too much information, 'to the point', direct	16
-1	LDJUG	juggling multiple on board systems felt unnatural when learning	16
0	NUNS	do not unsettle or make the driver anxious or vulnerable feeling	16
1	POSI	positive emotions arising from skill and control (relaxaing, invigorating, joy)	16
2	DEL	mention of delegating certain aspects of control with passenger if driver chooses	16
4	ROB	robust feeling or chunky	16
5	AEST	aestheticllly pleasant	16
5	COZY	sense of cosiness or homeliness in cabin	16
6	RAD	controls radiate out from steering wheel according to importance	16
9	COP	reference to the car being like a competent passenger or copilot	15
1	UNC	driving feels so natural it can be done on autopilot or unconscious	15
3	FROAD	feeling the road or wheels	15
3	NGAM	shuld not feel like playing a game or disconnected	15
4	SMOO	smoothness or fluid feeling action	15
4	NHOL	not hollow, tacky, plasticky or cheap, flimsy	15
4	GRI	grippy	14
7	VFB	some visible feedback is natural (flashing, change of position etc)	14
9	HML	refereences to friendly humanlike qualities	14
-1	LDFOR	Learner drivers surprised by strength of physical sensations or forces or the sensation of speed.	14
1	STRASK	strategic coordination/hgh level skills feel natural (route, overall mgmt, smoothness, competence)	13
4	PREC	feeligns of precision	13
4	MAT	preference for mattfinish	13
5	PRIV	car keeps appropriate thinsg private; concern about info going too widely	13
6	SWI	mechanical switches rather than digital clicks	13
8	NTMI	not too much info presented	12
8	NGAD	avoid gadgets or gimmicks	12
9	PERS	car is personified, anthropomorphed FRIENDLY RELATIONSHIP	12
10	EMER	Recognises emergencies or unsafe situations and helps out	12
11	VIN	expression of general voice control input preference; unclear which type	12
11	VINN	the car speaking does not feel natural	12
11	NLU	preference for the car having natural language understanding voice input	12
11	POL	car messages must be polite, don't overstep mark, respectful	11
11	UNAM	car messages should be unambiguous	11
0	PRED	general predictable and familiar behaviour by the car feels natural	11
2	IFC	driver informationally in full control	11
2	PCD	private customisable domain; personalisation of interface to enhance control	11
7	IFB	inherent or 'real world' feedback. You know it's doing it because you can see/hear/feel it actually	10
9	NTFR/NTFRIE	not too humanlike or friendly, it shouldbe a formal relationship	10
11	TIME	car messages must be well timed; timely	10

FIGURE 4.4 EXAMPLE OF THEMATIC CLUSTERING AND CONTENT ANALYSIS SPREADSHEET

4.8.4 INDEPENDENT CODING ANALYSIS

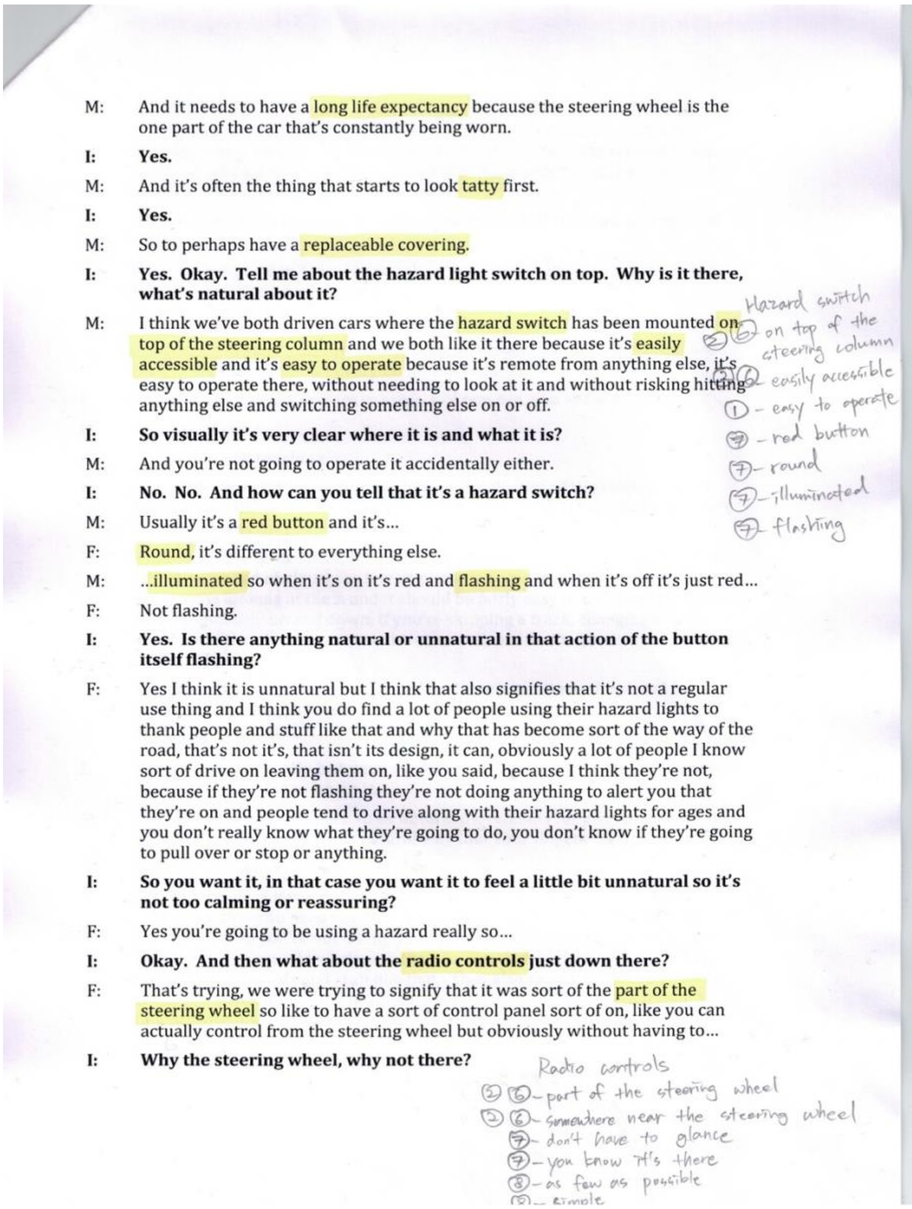
It is recommended in qualitative analysis to use independent coders (Braun and Clarke, 2006; Saldaña, 2015). A procedure presented as a 'pack' was therefore sent to three independent researchers. They were all automotive doctoral researchers at Brunel University, specifically one PhD researcher studying automobile-smartphone integration and automotive applications, one PhD researcher studying perception of automobile vibrations, and one PhD researcher studying driver-seat biometrics. They were considered suitable because two had undertaken qualitative narrative type research on drivers before using Thematic Analysis, while one had specialist knowledge of affective aspects of driving. The checking procedure required:

1. Free (eclectic) coding of three pages of randomly selected transcript from Sessions 3 and 4. Given the research question, the researchers were encouraged to create as many relevant codes as they felt appropriate and then write them on the transcript together with their rationale. Because of the length of each transcript exceeding three hours (60 pages) of verbal data, it was not feasible to ask for a whole transcript to be checked.

2. Using the codes the checkers identified in (1) above they were asked to start to group them into higher-level themes. The writing down of thought processes was encouraged.

3. Only after stages 1 and 2 were complete and returned, were participants emailed the tentative list of themes that had been developed from the results (i.e. the 'automotive ergonomic' themes). Each theme was concisely explained and numbered. Participants were then asked to code three different pages of transcript using this scheme by writing the number of the theme on the transcript in every place they thought that theme occurred. This was less time consuming for these independent researchers than coding from the master list of all 179 codes (the process used by the main researcher). This was to check the objectivity of mapping themes to transcripts.

Three code checkers were felt to be adequate for data deriving from ten participants. A final thematic *grouping* (clustering) check involved a master's level psychology research assistant (specialising in health and wellbeing research and measurement at Oxford Brookes University) and giving him all the codes, in the form of a card sorting exercise (Hanington and Martin, 2012). Although a plausible alternative grouping of themes was obtained, it did not offer clear advantages over the 'automotive ergonomic' framework. This is because each of those themes more richly and specifically described some characteristic of naturalness, suggesting 'valence', 'dominance' and 'activity' which are considered important in measurement of affective type 'feeling-based' constructs (Ekkekakis, 2013).



M: And it needs to have a long life expectancy because the steering wheel is the one part of the car that's constantly being worn.

I: Yes.

M: And it's often the thing that starts to look tatty first.

I: Yes.

M: So to perhaps have a replaceable covering.

I: Yes. Okay. Tell me about the hazard light switch on top. Why is it there, what's natural about it?

M: I think we've both driven cars where the hazard switch has been mounted on top of the steering column and we both like it there because it's easily accessible and it's easy to operate because it's remote from anything else, it's easy to operate there, without needing to look at it and without risking hitting anything else and switching something else on or off.

I: So visually it's very clear where it is and what it is?

M: And you're not going to operate it accidentally either.

I: No. No. And how can you tell that it's a hazard switch?

M: Usually it's a red button and it's...

F: Round, it's different to everything else.

M: ...illuminated so when it's on it's red and flashing and when it's off it's just red...

F: Not flashing.

I: Yes. Is there anything natural or unnatural in that action of the button itself flashing?

F: Yes I think it is unnatural but I think that also signifies that it's not a regular use thing and I think you do find a lot of people using their hazard lights to thank people and stuff like that and why that has become sort of the way of the road, that's not it's, that isn't its design, it can, obviously a lot of people I know sort of drive on leaving them on, like you said, because I think they're not, because if they're not flashing they're not doing anything to alert you that they're on and people tend to drive along with their hazard lights for ages and you don't really know what they're going to do, you don't know if they're going to pull over or stop or anything.

I: So you want it, in that case you want it to feel a little bit unnatural so it's not too calming or reassuring?

F: Yes you're going to be using a hazard really so...

I: Okay. And then what about the radio controls just down there?

F: That's trying, we were trying to signify that it was sort of the part of the steering wheel so like to have a sort of control panel sort of on, like you can actually control from the steering wheel but obviously without having to...

I: Why the steering wheel, why not there?

FIGURE 4.5 EXAMPLE OF INDEPENDENT CODER'S NOTES ON A TRANSCRIPT

4.9 RESULTS

Ten people took part in the workshops (mean age 41, six males and four females). They were mostly very experienced drivers (with few students as a result of the summer vacation timing) resulting in a broad age spread and more even gender balance than the previous study, which was considered advantageous. Seven were British nationals, two were South Korean and one was an Indian national studying in the UK. In summary the participants owned two 'small' cars, two 'family' cars, two 'medium-premium' cars, two 'large-premium/luxury' cars, one 'SUV', and one 'sports' car. Three hours was sufficient to complete the activities with each of the groups. Five sessions were conducted (the rationale for this was explained in Section 4.8.1).

4.9.1 *Results of the thematic analysis*

The five transcripts together provided 1770 relevant statements of the form 'X feels natural/unnatural [when using car secondary controls]'. Similar codes were combined. Prolific codes (more than 20 incidences, i.e. more than two per participant on average) were subdivided to make patterns more apparent (Patton, 1990). Rare codes that appeared only once were ignored. This gave a master coding list (Saldaña, 2015) of 179 naturalness codes. This is within the range suggested by Friese (2014). Each code was expressed on average 10 times across all workshops. Part of the master coding list is shown in Figure 4.4.

The initial attempt at clustering the codes into theme categories unintentionally used several 'automotive ergonomic' type categories (e.g. 'low visual demand') such as those found in automotive interface texts (e.g. Bhise, 2011; Harvey and Stanton, 2013) as well as some 'interaction design' stereotypes (e.g. 'physicality' and 'sentience'). Aware that this may have been prejudiced by the researchers' professional backgrounds, various attempts were made to group the 179 codes into *alternative* themes to maintain objectivity – consistent with 'bottom up' inductive research (Patton, 1990). To do this, all the codes were first printed and cut into paper strips as a 'sorting exercise' to explore alternative groupings and taxonomies (Hanington and Martin, 2012; Saldaña, 2015). By consulting academics in different departments across Brunel University linked by an interest in human-centred design (in computer science, social science, and psychology), and 'brainstorming' with the Oxford Brookes psychology researcher described in Section 4.8.4, four possible alternative taxonomies were developed. Firstly, categorising the 179 codes according to what *human need* was apparently served by each code gave three main themes: 'allows the driver to concentrate on driving', 'provides a satisfying game of control' and 'provides a mobile living space'; these were fairly evenly balanced with respect to code count and the categories appeared relatively internally and externally homogenous (Patton, 1990). Secondly, the '*sensory split'* approach, common in the social sciences (Degen, 2014) analysed what sensory channel each code related to; these were 'how it looks', 'how it feels', 'how it sounds', or 'what thought it demands'. However, these categories were highly unevenly balanced (with respect to frequency) making it unsuitable for the final thematic grouping. Thirdly, an attempt was made to organise the 179 codes into Jordan's (2002) framework for *'Design for Pleasure'* using its categories of 'Functionality', 'Usability' and 'Pleasure', and finally the Gaspar et al (2014) *'Driver Needs'* dimensions of 'Touch', 'Operation', 'Function', 'Sound', 'Styling', 'Location', 'Concept' and 'Novelty'. However, neither of these categorisations gave complete or satisfactory thematic groupings because about 30-40% of the codes were always 'left over' impossible to allocate. By contrast the initial clustering resulted in a greater number of distinct descriptive themes (13) of roughly equal frequency with no codes ‘left over’. It was therefore selected. Final coding was then carried out by marking the 179 final codes (as acronyms) on fresh paper transcripts.

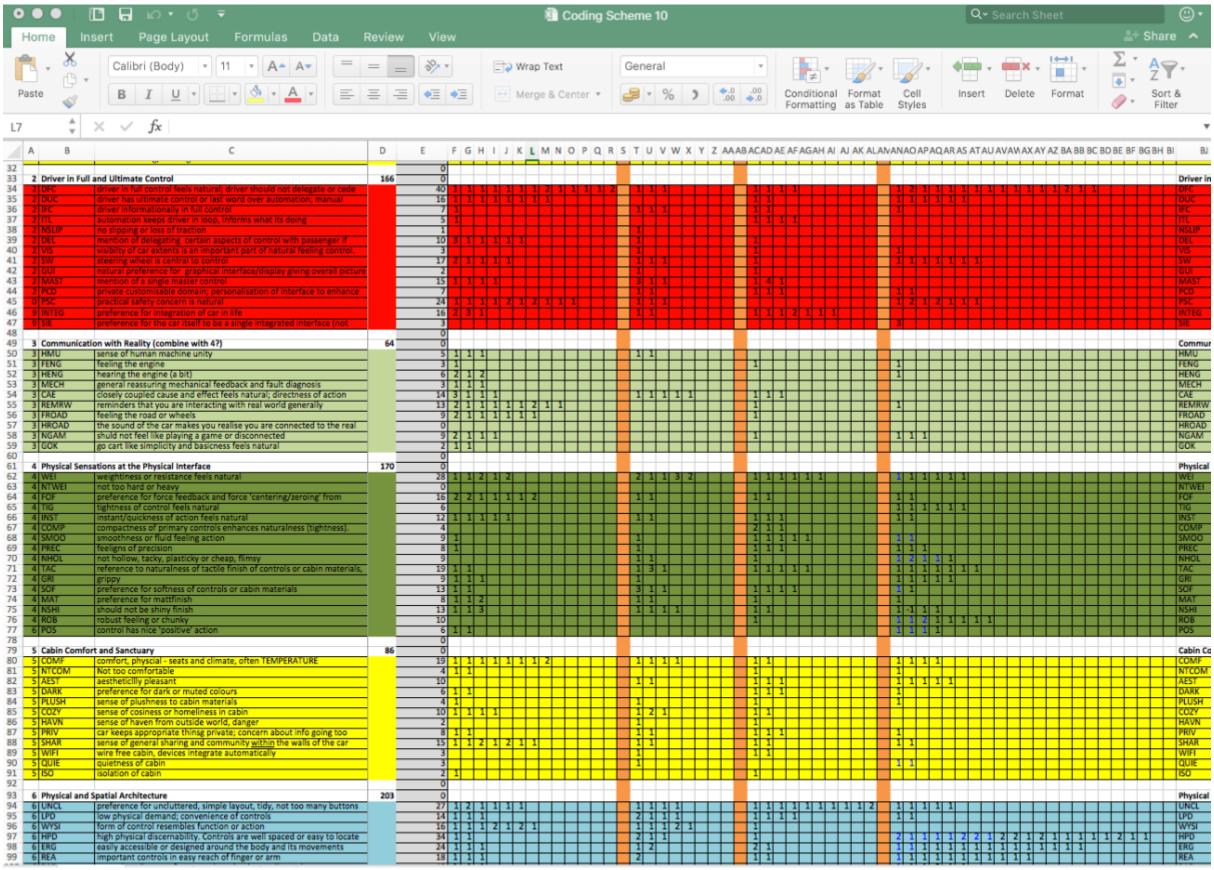


FIGURE 4.6 SCREENSHOT SHOWING A PART OF THE CODING SPREADSHEET OF 179 FINAL CODES, AFTER FINAL CODING BUT BEFORE FINAL CLUSTERING

In grounded theory analysis, the dataset must be allowed to ‘speak for itself’, and it cannot be supposed that the phenomena under investigation will naturally group into ‘convenient’ themes of equal frequency (Glaser and Strauss, 2009). However, a *desirable* quality of any thematic framework is a manageable number of highly descriptive themes each representing a similar proportion of responses. Fortunately, the 179 codes clustered logically into the 13 ‘automotive ergonomic’ themes (described in Section 4.10) of roughly equal weight in terms of both their *number* of constituent codes and their overall *frequency* of occurrence. One theme was disregarded because it related only to the sensitisation question (about ‘the sensory experience of first driving experiences’) which was not in itself a research question. Another theme, ‘positive feelings arising from full control’ was ignored because it was considered by the research team and independent code checkers to be only a subset of (or consequence of) a related theme about being in full control. The final themes are listed in Section 4.10.

4.9.2 Results of the Content Analysis

Table 4.3 shows the 15 most frequent naturalness codes (i.e. repeated semantic statements of similar meaning, relating to naturalness of car secondary controls). They are presented in descending order as a proxy for salience.

Table 4.3: The 15 most frequent 'naturalness' codes in the data, most frequent at top

Rank	*Naturalness-related code (ie automobile characteristic or perception)*
1.	Driver being in full control feels natural; driver should not delegate or cede control
2.	Natural-feeling interactions should not distract from the primary driving task
3.	Minimum utility while in motion feels natural; not too much input or adjustment required
4.	Natural-feeling controls are physically 'discernible' (well-spaced, locatable by touch)
5.	Old fashioned controls feel natural in an automobile; skeuomorphism preferred
6.	Low visual demand (generally) feels natural
7.	The control being intuitive to use feels natural
8.	General sense of ease or simplicity feels natural, relaxed feeling while driving
9.	Weightiness or physical resistance of controls feels natural
10.	Naturalness preference for an uncluttered, simple layout, tidy, not too many buttons
11.	Not too much choice or too many decisions feels natural
12.	Not easy to make unintended inputs, false alarms etc., feels natural
13.	Natural-feeling controls have familiar, fixed, predictable locations, mapping and layouts
14.	A natural-feeling control's action is obvious
15.	Practical safety concern – drivers are naturally aware of factors that might lead to danger

4.9.1 Observations

While the data was intended for only Thematic and Content Analysis, described below, certain observations were considered worthy of separate reporting because they were relevant to the research question, or demonstrated the potential of using research methods which are still novel in the automotive interface design sector.

1. In the **sensitising exercise** people shared often difficult or embarrassing 'first driving' memories, and this helped create the desired interactive workshop environment with an atmosphere of openness. Participants often reported very detailed sensory memories of the way their automobile's controls felt *the first time* they used them (perhaps twenty or thirty years before). This perhaps indicates the significance of the research topic – that interaction

with cars' controls may be a deeply sensory 'life experience'. It is hard to imagine such strong sensory memories deriving from someone's first use of a toaster, for example.

2. The '**Think Aloud' sessions** elicited much relevant data. Perhaps because participants had never used these particular car controls before, or because they were removed from their normal context of a car cabin, participants' perceptions and expectations were reported in great detail, as compared to descriptions of their *own* cars' controls in some of the pilot tests.

3. Much activity and animated creative reflection was observed in the **'Natural dashboard'** arrangement task. Drivers' explanations of their choice of components, materials, and overall layout were highly relevant to the research question. The 'unpacking' of sensory and material choices also resulted in valuable data. *Participants' 'natural-feeling dashboard' creations tended to be sparse, simple, convenient, assistive, with large tactile controls (e.g. swivel vents) and predominantly matt and dark textures* (see Figure 4.7 for an example).

FIGURE 4.7: AN EXAMPLE 'NATURAL-FEELING DASHBOARD' CREATION FROM ONE WORKSHOP

4. In the **'Unnatural dashboard'** exercise the most activity and pleasure was observed of all the sessions, and several participants commented that it was easier to specify what they *disliked* about automotive interactions and what aspects and situations felt *unnatural*. By enquiring as to the meaning of 'unnatural' in this breaching exercise, an additional source of naturalness characteristics was captured. This maximised data capture. *Participants' 'unnatural-feeling dashboard' creations tended to feature small buttons (e.g. from calculators or 1980s radios), overly complicated settings (e.g. a window control that required dialling in an exact opening percentage numerically), unnecessary alphanumeric readouts (e.g. cabin temperature), loose wires, rough or metallic/shiny textures, lack of tactile or mechanical controls, and distracting feedback (e.g. bright flashes).* See Figure 4.8 for an example.

5. The activity-oriented workshops apparently succeeded in **engaging participants** and maintaining their interest to the end of what were rather lengthy sessions. Almost all participants commented during or afterwards that they had enjoyed it, unprompted, and many declined to take the incentive voucher as a result.

FIGURE 4.8: AN EXAMPLE 'UNNATURAL-FEELING DASHBOARD' CREATION FROM ONE WORKSHOP

4.10 NATURALNESS FRAMEWORK WHICH EMERGED FROM THE EXPLORATORY DESIGN WORKSHOPS

1. Familiarity and Predictability

Participants' responses suggested that secondary controls (and responses) which are familiar, recognisable, predictable, not alarming, and safe, tend to feel natural. It was also suggested that this 'familiarity' can be learned and becomes natural-feeling over time.

2. Driver in Full and Ultimate Control

Participants' responses suggested that secondary controls and systems that make the driver feel fully in control, tend to feel natural. The driver should always be 'in the loop' and ultimately in control even if executive control is sometimes delegated. Some data suggested that arranging controls around steering wheel or master display may help create this feeling.

3. Communication with Reality

Participants' responses suggested that It feels natural for an automobile to communicate certain 'real-world' information about the road, its mechanicals and environment. This appeared to be a 'reminder' that the driving interaction is partly an interaction *with the real world* and not a game or inconsequential human-computer interaction.

4. Weighty Physical Sensations

Participants' responses suggested that certain physical sensations and perceptions at the interface, mainly felt through the hands, feel natural in secondary controls. Examples are heaviness/weight (rather than light feeling), tightness (rather than loose feeling), directness, precision, robustness, and 'tactility' (rather than hard or shiny feeling).

5. Cabin Comfort and Sanctuary

Participants' responses suggested that a comfortable, private, dark, protected, relaxing, homely, aesthetically pleasing cabin with good visibility seemed to be natural-feeling

6. Uncluttered Cabin Architecture

Participants' responses strongly suggested that a natural-feeling cabin layout features secondary controls which are uncluttered and efficiently located, easily locatable either by virtue of their visual distinctiveness or by touch alone (when drivers' visual attention is on the road). This makes accidental or unintended inputs rare. The data mildly suggested that rotary dials and mechanical switches may feel more natural than digital clicks in this regard.

7. Low Visual Demand

Participants' responses strongly suggested that natural driver-car interaction with secondary controls demands little visual attention away from the primary driving task. Non-visual modalities were suggested for feedback. The data suggested many natural-feeling controls may be operated without looking. Analogue dials may be more natural than digital displays.

8. Low Cognitive Demand

Participants' responses strongly suggested that that natural-feeling interaction does not cause cognitive distraction from the primary driving task. Minimal information, choices or concerns should be presented to the driver when in motion. The data further suggested that secondary control shape and action should be well-mapped to its function and response; control actions should be obvious or, at worst, clearly labelled.

9. Humanlike Driver-Automobile Partnership

Participants' responses suggested that interaction with Intelligent automobile features or automation will feel more natural if the automobile behaves as, and is perceived as, a helpful co-driver. Examples of this were 'informative', 'polite', 'helpful' and 'proactive' behaviour.

10. Humanlike Sentience and Learning

Participants' responses suggested that an intelligent automobile would feel more natural if it sensed, processed and understood things in a humanlike way, such as remembering preferences, predicting events, adapting to situations and being empathetic (displaying social and emotional awareness).

11. Humanlike Verbal-Auditory Communication

Participants' responses suggested that an intelligent automobile could most likely be interacted with by voice. Data suggested that naturalness will be enhanced by perfect natural language understanding, the car 'speaking only when spoken to' and keeping messages brief, timely, polite, concise and unambiguous. The tone of voice would be neither too humanlike nor too machinelike.

Numerically the most frequent three themes were all usability issues – 'Low Cognitive Demand' followed by 'Uncluttered Cabin Architecture' and 'Low Visual Demand'. The results may best be expressed by the following framework (Figure 4.9) because it logically groups the 11 themes into four higher order categories: 'familiarity and predictability', 'physical themes', 'usability themes', and 'future intelligence' themes.

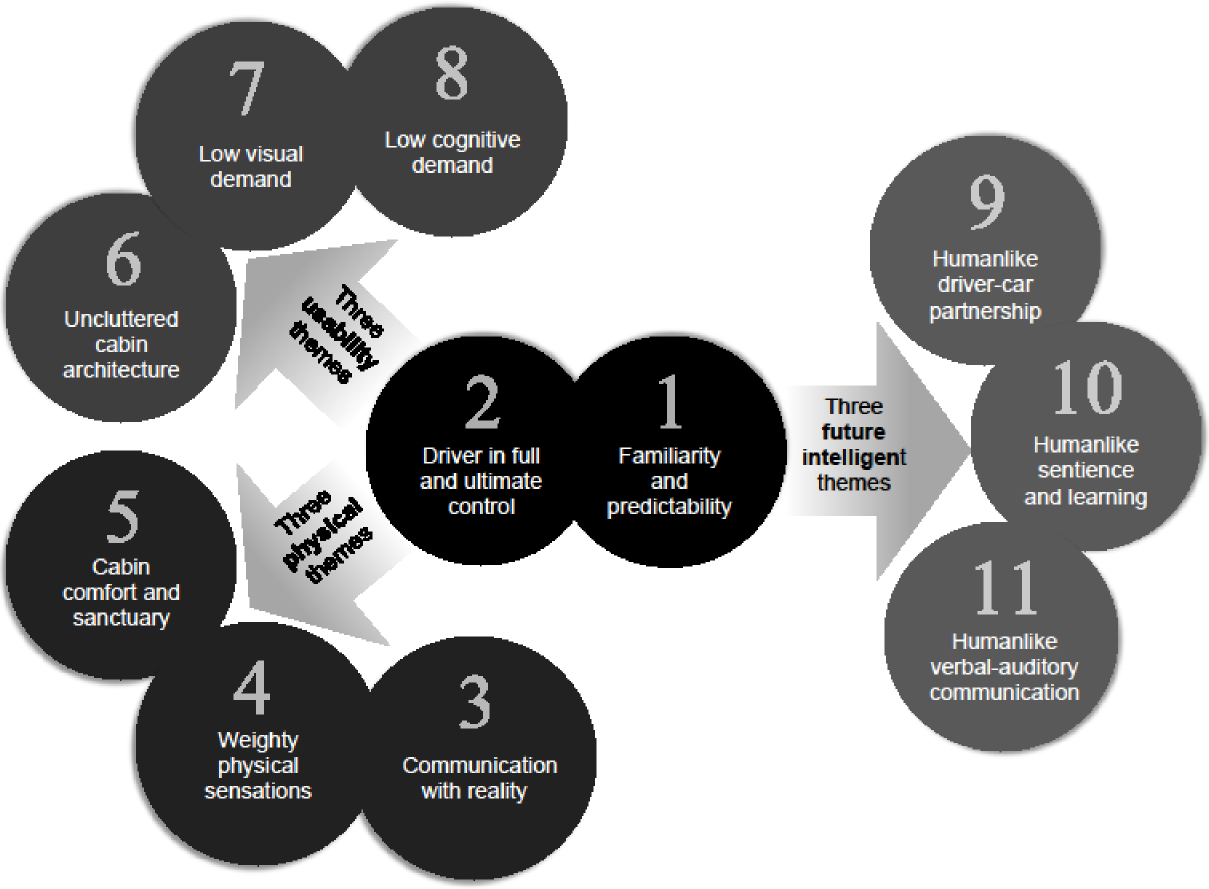


FIGURE 4.9: THE 11-THEMED FRAMEWORK OF DRIVER-CAR NATURALNESS DETERMINED FROM THE EXPLORATORY DESIGN WORKSHOPS USING THEMATIC ANALYSIS

4.11 DISCUSSION

The mixture of methods, exercises, modalities and task locations provided 1770 statements directly related to the research question. This exceeded the expectations based on the experience of the previous study. This may have been a result of the greater length of total 'contact time' with participants, the greater variety of activities, or the social and negotiated aspects, or simply because the questions were more direct and explicit in asking about 'naturalness' than in the previous study.

Many of the themes found have strong parallels with those found in the framework developed in Chapter 3. Being independently derived, there are inevitably some differences in theme meanings, distinctions and emphasis between the two frameworks. This was likely to have been the result of the different participants, different elicitation methods and different research designs used. A full discussion of these differences may be found in the chapter concerning triangulation (Chapter 6). Briefly, Themes 2, 3, 4 and 5 have clear parallels with Theme 1, 2, 3 and 4 in the previous study's framework. The usability themes (Themes 6, 7 and 8) were more frequent and more dominant in this study compared to the previous study (only Theme 5) perhaps because of the use of unfamiliar controls. Usability therefore 'expanded' from one rather general theme in the previous framework to three much more specific themes concerning the 'physical', 'visual' and 'cognitive' aspects of usability – a distinction familiar to ergonomists. The theme from the previous study about the natural-feeling intelligent car behaving as a 'single intelligent being' did not arise in this study at all. The three 'future intelligent themes' in the present framework are roughly equivalent to the 'social and intelligent' themes in the previous framework, with subtly different distinctions.

Considering the practical implications of all the 11 themes, the data suggested that general **familiarity and predictability** and **being in full control** of an automobile's controls are essential to *any* automotive interaction, but here there was evidence to suggest that these aspects contribute to natural-feeling interaction also. Familiarity and control are physical design characteristics which may perhaps elevate 'natural' automotive interface design above the traditional 'usability' approach. The two themes however appear at odds with many of the novel interface arrangements proposed for future and self-driving cars. Skeuomorphism, whereby novel interfaces imitate older more familiar controls and interaction stereotypes, is a recognised approach in interaction design (Shedroff and Noessel, 2012) which maintains 'familiarity' and draws upon existing metaphors. Such an approach may be necessary for novel features in automobiles to be perceived as natural. A 'Haptic Shared Control' approach to self-driving cars (essentially a moving controller shared by the automation and the driver; Abbink et al, 2012) may also maintain a semblance of full control and predictability.

The ‘**physical’ themes** suggest that ‘feel’ and some semblance of physical connection are also important for natural-feeling secondary controls, which again may be challenging in a self-driving car. One or more physical controls with weighted precision in their operation may need to be retained even when there ceases to be any *mechanical* connection between interface and car. This is rather at odds with the emergence of the ‘glass cockpit’ (Harvey, and Stanton, 2013) and manufacturers’ visions of automotive futures (e.g. Fowler, 2015) which often show no physical controls at all, other than steering wheel and pedals. Ford’s customer experience with its *MyTouch* infotainment system, which was criticised by customers and media alike, also suggested that entirely digital controls may not feel entirely natural (Lanks, 2015). Ford’s reaction was in fact to reintroduce some physical knobs. Interestingly the very latest premium car dashboards which are largely touchscreen, tend to retain manual round dials for volume and temperature. The role of comfort is less clear. It may be that some participants’ *human* needs for basic comfort and privacy were misinterpreted as a naturalness requirement. While many participants stated that natural interaction had to be physically comfortable, several participants commented that a natural-feeling automobile would not be *too* comfortable (because that would make them less alert). Perhaps ‘comfort’ in natural interaction simply means meeting minimum ‘ergonomic’ standards for reach, posture and operation. Its inclusion in the framework is tentative but it cannot be ignored because of the volume of associated data.

The most frequently expressed themes were the three **usability themes**. This salience was perhaps because drivers in the workshops were operating controls that were new to them, and using an unfamiliar dashboard and car. Usability aspects were however *so* prominent in perceptions of natural-feeling interaction, that combining them into a single category seemed inappropriate. The breaching exercise regarding ‘unnatural’ aspects of dashboard design, helps to explain why the usability themes are expressed as the ‘lack of a negative quality’ – e.g. ‘uncluttered’. These names were left unaltered because participants appeared to prefer them and in practice it was hard to find ‘positively phrased’ equivalents. For example, participants readily expressed and engaged with the idea of natural-feeling automobiles being ‘uncluttered’ whereas corresponding positive equivalents such as ‘spaced-out’ tended to change the meaning and were not well understood or received.

The final three themes represent **humanlike intelligence** and suggest how natural-feeling interaction might plausibly evolve when automobiles have greater intelligence and agency than their drivers, echoing the recommendation in Cooper et al (2012) that technology should behave as ‘a likeable person’. It suggests natural-feeling ‘intelligent’ cars will need to be ‘sentient’, humanlike, helpful but subservient. Inevitably the inclusion of a voice-based activity will have prejudiced answers towards verbal-auditory communication, but this theme frequently arose spontaneously *before* that exercise, so it was included in the framework.

Comparing the findings to the **literature on naturalness** (Section 2.6) suggests many predictable parallels but also some novel contributions to knowledge about what specifically makes *driver-car* interaction feel natural. Considering each of the 11 themes from the results framework, Theme 1 (familiarity) could be interpreted as equivalent to the naturalness of established cultural HCI stereotypes (as in Malizia and Bellucci, 2012) or equivalent to the ‘well-practised response generation’ described by Goodrich and Olsen (2003). However, the associated data also implied that perceptions of unsafety, surprise or alarm are *un*natural (and that natural interaction feels safe and unsurprising) – which goes beyond conventional academic interpretations of the term ‘natural interaction’. Predictability of metaphor is often considered important to ‘natural’ interaction (O’Hara et al, 2012; Celentano and Dubois, 2014) and this may be related to familiarity. Control (Theme 2) has not generally been considered fundamental to naturalness (unless the ‘natural’ inclination of the human being is regarded as ‘control’ or mastery of his surroundings) but these findings suggest it might be, in the case of automobiles. Theme 3 (reality) is again referenced in the ‘natural’ reality-based interaction described by Hornecker (2011) and Jacob et al (2008); Theme 4 (rich skilled physicality) hints at the ‘naturalness’ of naïve/causal physics described by those same authors but qualifies it by explicitly suggesting qualities of weightiness, tightness and springiness (generally regarded in the literature as contributors to interaction pleasantness or satisfaction but not naturalness). Comfort (Theme 5) is the only finding not seen at all in the literature (again, unless an anthropological view is taken that all humans naturally avoid *dis*comfort). Themes 6, 7 and 8 (the usability themes) essentially concern low visual and cognitive demands in naturalness and these may be related to sensory motor skill transfer naturalness or cognitive skill transfer naturalness (described by Berard and Rochet-Capellan, 2015). These ‘usability’ qualities may also be seen as a secondary benefit of adopting a ‘natural’ RBI or tangible approach (Dourish, 2004) – for example a physical round volume knob on the dashboard probably has lower visual and cognitive demands than an alphanumerical volume selector with ‘up’ and ‘down’ keys, especially if it is buried in an infotainment menu. In the literature usability is sometimes assumed to lead to naturalness, but the relationship may be more complex (O’Hara et al 2012) for example too much ‘naturalness’ may sometimes *hinder* usability because the point of computing is to be more powerful than natural (Pieraccini and Huerta, 2005) or it may be naturalness that leads to perceptions of usability rather than the other way around (e.g. Susini et al, 2012). Finally Themes 9, 10 and 11 (the ‘humanlike’ themes associated mainly with in intelligent cars) possibly provide a novel contribution to understanding driver-car naturalness. The metaphor of ‘partnership’ is the closest finding to the future-focused proposals of Flemisch et al (2010) who suggest automated cars be designed to feel more natural by using a ‘horse-rider’ cooperation metaphor varying between ‘loose rein’ or ‘tight rein’ control depending on scenario. These three themes together suggest that intelligent car interfaces designed to empathise, listen, learn, predict and speak, may feel more natural to use than car interfaces that do not. Such humanlike qualities have generally only been associated with human-*robot* naturalness (Van Dam, 1997). As in the previous chapter, this is the main contribution of this study to knowledge about naturalness – and it may be applicable to other intelligent applications beyond just the car.

Although the thematic groupings were perhaps influenced by the researchers’ knowledge of automotive interface design and interaction design, it was felt on balance that this was a valid approach, given that the end users of any derived rating scale would be automotive designers and interaction designers. The final number of 11 themes was also felt to be manageable for busy professionals to use as a ‘heuristic’. It is suggested that an automobile, system or control which complies with as many of the themes as possible will be perceived as more ‘natural’ than one that does not.

Not all themes will be applicable to every system. Some of them may oppose each other – for example too many physical controls (Theme 4) may lead to a cluttered cabin (opposing Theme 6), and too much feedback from the road and car (Theme 3) may undermine comfort (opposing Theme 5). Too little cognitive demand (Theme 8) may lead to boredom and lack of attention – probably unsuitable for any vehicle interface design. This suggests that the themes are not always orthogonal, and sometimes inversely proportional. It also suggests there may be optimum ‘mid-range’ levels for some aspects of naturalness – with too little or too much of any one quality detracting from the ‘natural feeling’.

4.12 Limitations

One limitation on the data was not capturing participant demographics. Because this was the only study not to take place wholly inside a real car, the fixed venue (the university laboratory) may have attracted younger and university educated participants compared to the population as a whole, though the vacation timing ensured more postgraduate students and administrative staff took part than undergraduates or academics. Such participants may be more computer literate than the general public and this may have affected the results. They may have had some knowledge about Brunel Design and may have socially-mediated their answers to suit. It was not possible to hide all the branding on the Jaguar test car (in any case the overall shape and interior are highly indicative of its manufacturer and age) and despite hiding branding the loose switchgear, some of it was identifiable where people had experience of the particular car it came from. This may have prejudiced some of the results. Although the word ‘natural’ was used explicitly in most questions, and an attempt was made to define what sense it was meant in, participants may have unknowingly conducted exercises according to what they believed to be *good* design rather than *natural-feeling* design. One third of the participants learned to drive in countries where the driver sits on the left-hand side of the car, a greater proportion than in the other studies. The test car and co-creation templates were both right hand drive which might have been disorientating for those participants. In retrospect those templates could have been inverted as required. It is not known what effect driving-side conventions has on naturalness perceptions (again, handedness was not tested) although at best it might make these findings more generalizable to other countries and cultures. The payment of an incentive was thought necessary to attract non-car enthusiasts, but the degree to which this was successful was not tested. Payment of incentives can also prejudice results. Finally, the experimental set up and physical habitat in the laboratory may have biased results. For example, the presence of large numbers of small controls on a table (Figure 4.2) may have felt overwhelming and contributed to the finding that natural-feeling interfaces should be *uncluttered*; similarly, the fact that automotive controls can actually appear rather lightweight and flimsy when removed from a car might have biased the results regarding natural-feeling controls being ‘weighty’ and ‘robust’.

4.13 CONCLUSIONS

In order to explore the characteristics of natural-feeling interaction between automobile drivers and their secondary controls, five activity-based exploratory design workshops were conducted using a combination of six different artefact-focused methods to elicit the perceptions of ten ordinary car drivers. The data was grouped into 11 relatively homogenous themes after thematic analysis and independent code checking. In summary the data suggested that natural-feeling secondary car controls should appear familiar, behave predictably, and allow drivers’ attention to remain on the road through established principles of ‘usability’. Retention of some physical controls and the perception of being ‘connected to the real world’ might also feel natural. A future intelligent automobile might still feel natural if it can behave as a helpful, learning, listening, speaking co-driver.

CHAPTER 5

ETHNOGRAPHIC OBSERVATIONS OF DRIVERS INTERACTING 'NATURALLY' WITH SECONDARY CONTROLS

5.1 AIM

In previous chapters, Contextual Inquiry and exploratory design workshops gave two possible frameworks for what constitutes and creates natural-feeling interaction between drivers and their secondary controls. At this point, an ecologically valid **observational** study of how drivers interact 'naturally' with their **present-day secondary controls on real journeys** was considered necessary, in accordance with human centred exploratory design research practice (Giacomin, 2012b). This was to avoid overreliance on narrative methods and methods conducted away from realistic on-road scenarios.

The first study had suggested that drivers had rather clear and consistent impressions of **natural-feeling car controls in the future.** However**,** it was doubtful that the limited future scenarios given to participants previously had adequately explored and recorded such perceptions. As discussed in Chapter 1, dashboards and secondary system interactions appear to be evolving rapidly, yet automotive development lead times can still be very long. Secondary systems designed today might only appear in production vehicles in three or four years' time. It was therefore decided to attempt to make this research's findings more relevant to the design of likely future cars as well as present day cars, by conducting **observations of drivers interacting with some simulated 'intelligent' secondary controls and systems of the (possible) future**. Realistic and convincing simulations can help participants imagine and explore futuristic interactions which they might have trouble imagining unsupported (Ylirisku and Buur, 2007; Dunne, 1999).

The aim of conducting the two observational studies (one 'current day' and one 'futuristic') was primarily to check that the two frameworks described naturalness as experienced on *real-life* journeys; secondly to resolve the uncertainties and conflicts in the two proposed frameworks perhaps clarifying the discrepancies between them particularly with respect to the 'social and intelligent' and 'future' themes; and thirdly to suggest amendments to improve their likely 'future relevance'. Ideally a *third* framework would be developed from the data, for the purposes of method triangulation (Decrop, 1999). The overall research question remained the same, but the objectives needed to be somewhat more specific as explained above. Five research objectives were devised with the research director:

1. How do drivers appear to ‘naturally’ or ‘instinctively’ interact with secondary controls in real-world scenarios during road journeys? What aspects appear natural and unnatural?
2. Do drivers appear to ‘naturally’ operate their secondary controls fully consciously, semi-consciously or unconsciously when managing the primary driving task on road journeys?
3. Does secondary control weight, shape and feel, and driver comfort, appear to be essential to natural-feeling interactions or are they more general ‘expectations’?
4. How significant do ‘usability’ factors (uncluttered controls, low visual demand, low mental demand) appear to be to un/natural-feeling interaction during real road journeys?
5. What characterises un/natural-feeling ‘sentient’ and ‘verbal-auditory’ interactions with intelligent secondary systems in (simulated) intelligent future cars on road journeys?

5.2 Research methods

Three main methodologies emerged in the literature review which showed potential for observational studies inside a moving car. All were drawn from the field of human centred design (Giacomin, 2012a) using the compilation of design research methods found in Martin and Hanington (2012) as a starting point. All methods were ranked using criteria of:

1. Potential to answer the research questions
2. Experimental viability inside a confined moving car cabin on public roads
3. Potential for understanding private, silent, interactions between driver and car
4. Low safety/distraction risks to driver (and researcher)
5. High ecological validity (‘natural’ interactions logically need to be studied in ‘naturalistic’ settings rather than artificial or contrived settings).

This left three possible methods, one being largely a subset of another, described below.

5.2.1 Ethnography

Ethnography, the scientific observation and description of peoples and cultures ‘at work’ with their associated practices, customs and habits (Wolcott, 1999, Dourish, 2004), is usually considered an essential component of qualitative, human-centred exploratory research (Giacomin, 2014). A central tenet of ethnography is to observe people ‘in the wild’ doing their everyday activities, not in “highly constructed situations” (Helander, 2014). Flexible research design is normal in ethnography and studies typically take shape as work proceeds and unexpected events are followed up (Helander, 2014). Ethnographic methods have been used within the field of technology studies, to study the interactional ‘hot spots’ of a setting

(Cycil, 2016) either with the researcher physically present, or absent (using video). Ethnography in itself is a discipline comprising several different methods. However, the literature survey and ranking exercise strongly suggested that only one, Participant Observation, would be suitable for studying driver-car interaction.

5.2.2 Participant Observation

Participant observation, a mainstay of ethnographic information collection which has been used for over 100 years (Kawulich, 2005), is the "systematic description of events, behaviours and artefacts in the social setting chosen for the study" (Marshall and Rossman, 1989 p78). This is usually achieved by means of 'active looking', informal interviewing and detailed field notes (Kawulich, 2005). The researcher may participate in the activity but in a rather limited way, maintaining impartiality and objectivity as befits an ethnographic observer; the researcher aims to establish rapport so that the individuals under observation simply 'act naturally' (ibid). Thereafter the researcher departs the setting and immerses him/herself in the data to try to understand what is 'really occurring' (ibid). Participant observation is claimed to develop holistic understanding of the phenomena of interest and its context which is both objective and accurate (Dewalt and Dewalt, 2002).

Only two types of participant observation were considered feasible in the present research, as defined by the theoretical stances of Gold (1958) – namely 'observer as participant' (which might be applicable to a researcher travelling in a car being driven by a participant, taking notes) and 'complete observer' where the researcher is hidden from view (as might be the case if a one-way video link was used to relay drivers' interactions to a laboratory). Spradley (1980) defines these as 'passive participation' and 'non-participation' respectively.

'Participant observation' best describes the ethnographic automotive studies by Meschtscherjakov et al (2011, 2015), Laurier (2005), Laurier and Philo (1998), and Cycil (2016). These researchers all studied real (driver-instigated) journeys on public roads, and occasionally used photography (non-flash) or video recording but most data capture was written in note books. Meschtscherjakov et al (2011) outline the challenges of using researcher-present observation in the cramped, private, busy, noisy and potentially dangerous moving car cabin; however they suggest that being present in the moving car allows the researcher to observe and react to important unexpected events. Meschtscherjakov et al (2011) suggested that drivers generally made the travelling researchers feel welcome in their cars, but that when observing driver-passenger interaction (the focus of their study) there was "rather artificial behaviour" for about the first 30 minutes. Behaviour was presumably considered to be 'natural' after that. To minimise bias, taking 'focused observation' field notes is often recommended, separating out observations from interpretations. The practice of *researcher reflexivity* (Kawulich, 2005) is also recommended for bias reduction. This means 'self-reflecting' on the position of the researcher and what effect this might have on the activity, observations and interpretation.

5.2.3 FUTURE FICTIONS AND THE 'WIZARD OF OZ' FORMAT

Future fiction, sometimes known as science fiction prototyping, is the use of science fiction story, film or comic, typically based on real science and technology, to explore the real world implications and uses of future technologies (Johnson, 2011). Ylirisku and Buur (2007) and Dunne (1999) suggest creating realistic but compelling scenarios to display a future that people may have difficulty imagining unsupported, using video or semi-functional prototypes.

So-called 'Wizard of Oz' formats are increasingly used in interface design research, to demonstrate highly intelligent systems where it would be unfeasible (technically, financially or logistically) to develop convincing functional prototypes (Dahlbäck et al, 1993). The format consists of a real person (the 'Wizard') concealed behind a curtain or remote video/audio link, who plays the role of the 'intelligent system' and interacts with a subject via speech, text or evident machine action. Usually the participant is not told that it was a *human* operator they were in fact interacting with, until the end of the experiment.

'Wizard of Oz' formats have occasionally been used in automotive interface research to simulate automotive interfaces which recognise emotion or affect (Eyben et al, 2010) or to simulate the interaction design of handovers in automated driving (on public roads) using a car with dual-sided driving controls. This format enables both the 'car' (i.e. the researcher acting as 'Wizard') and the participant to physically control the car at different times, and thereby simulate handovers, with the 'Wizard' (seated in the front passenger seat) hidden from view by a screen or curtain to maintain the illusion (Meschtscherjakov et al, 2016).

5.2.4 FACIAL CODING ANALYSIS

Some research has suggested that many basic human emotions present fairly consistently in the form of universal facial expressions that are common across ethnicities, nationalities and cultures. Ekman (1970) suggested seven common emotions happiness, surprise, fear, anger, disgust, sadness and neutral. Further analysis has suggested that there are 46 anatomically distinct facial expressions known as Facial Action Units (Ekman and Friesen, 1978) associated with these seven common emotions. Most commonly used in market research and advertising, Facial Coding Analysis is used as by proxy to gauge the moment-by-moment emotional and cognitive states of subjects (ibid).

5.3 STUDY OBJECTIVES

Two complementary participant observation studies were judged to be required:

- Study A: To answer Research Questions 1 to 4, a sample of ordinary drivers would be observed while driving long round-trip journeys on public roads incorporating a variety of road conditions including stationary traffic, urban driving and motorway driving.
- Study B: To answer Research Question 5, and to increase the future relevance of Research Questions 1 to 4, it was decided a speech interface for controlling intelligent secondary (mainly infotainment) controls would be prototyped using a 'Wizard of Oz' protocol. A sample of ordinary drivers would be observed interacting with it while driving on public roads.

In both studies the objective was to observe and record any behaviours, instincts and perceptions possibly related to natural-feeling interaction (and unnatural-feeling interaction) with secondary controls. This would be done by monitoring drivers' physical actions, body language, facial expressions, errors and any relevant verbalisations. Occasionally questions would be asked. Approximately the same *total* journey times were sought for Study A as Study B. Facial Expression analysis is time consuming to learn professionally, and highly subjective, but the researcher achieved a basic grounding adequate for the experiment.

The dependent variables of interest monitored were:

1. Positive facial expressions (delight/pleasure/relief/neutral/positive surprise) and negative expressions (such as negative surprise/fear/anger/disgust/sadness) were monitored. Whilst it was not assumed that naturalness is equivalent to positive emotion, or unnaturalness equivalent to negative emotions, it was assumed there would be a positive correlation from the literature review. On detecting a positive or negative expression there would then be a short opportunity (around 10 seconds) to explore this emotion 'in the moment' and ask about un/naturalness.
2. Body language indicating if an interaction felt natural or unnatural (largely a subjective judgment, but the driver's feelings could be explored verbally by questioning if required and it was safe to do so). In practice body language was closely linked to facial expression and monitored concurrently.
3. Drivers apparent intent/action and success in meeting goals via their car and its controls – i.e. whether operation of the car and its controls was successful or unsuccessful at meeting the driver's immediate need or longer-term goal.
4. Recording any spontaneous verbalisations and exploring these if it was safe (drivers were told that the study was about interactions between drivers and their cars, but not naturalness, to avoid prejudicing or biasing any verbalisations).

5.4 Key methodological decisions

Participant observation was considered the most suitable of the ethnographic methods for both Study A and Study B. For the observations to be ecologically valid (Patton, 1990) it was considered essential that the studies involved real journeys in real world road and traffic conditions. It was decided most technically feasible for the researcher to travel inside the car (therefore becoming part of the 'social setting' of the car) but taking only written notes.

Computerised 'natural language personal assistants' are increasingly common. They aim to help users with a wide variety of tasks using mostly natural language inputs without many specific 'command' words needing to be memorised. They may combine offline functionality for simple information retrieval with online abilities for researching or calculating more complex information. Such systems have frequently been suggested as a safer alternative to screen or keyboard interactions in the motor car (Weng et al, 2006). Furthermore, the study described in Chapter 3 had suggested that many drivers *expect* to be able to talk to their cars using natural language, as a 'natural' means of controlling its systems in the future. Therefore, it appeared optimal to simulate an in-dashboard natural language assistant which controlled car infotainment and some secondary controls, for the observations in Study B.

5.5 Study design

A pilot test suggested the researcher-observer needed to be either seated in the front passenger seat (to get an adequate view of the driver's face, eye movements, body and hand actions) or remotely observing via video link with good quality face video camera and ideally an upper body camera which ideally included a view of the dashboard, and if possible a third camera with the drivers forward facing view of the road (for context). This is complex but has become standard practice in (rare) naturalistic on-road studies (e.g. Guo and Fang, 2013) helping the observer to judge the driver's emotions, mood and intent. Facial expression interpretation and body language observation are two possible human centred design tools to help interpret perceptions and feelings which people may not be consciously aware of, nor perhaps capable of expressing otherwise (Giacomin, 2012b; Giacomin, 2014).

5.5.1 Study A: Drivers interacting with ordinary cars

It was decided that all the observed journeys should be 'driver instigated' for maximum ecological validity (i.e. they would have taken place anyway) and that the researcher should travel in the front passenger seat and simply silently observe interactions and take notes, in order to avoid distracting the driver. The journeys needed to be more than 30 minutes to gather enough data and to maximise the chance of drivers behaving in a 'natural' way

(Meschtscherjakov et al, 2011). One hour was considered optimal from the ethnographic pilot study conducted at the beginning of the first study.

5.5.2 Study B: Drivers interacting with simulated future cars

Devising a 'Wizard of Oz' type simulated intelligent 'natural language personal assistant' system for the futuristic participant observation had the intrinsic advantage that drivers would be forced to verbalise what would otherwise be private silent interactions, with clear benefits for the present research. Perceptions, errors, and nuanced transactional qualities might be observed and interpreted from drivers' words, tone of voice, body language and facial expressions. An opportunity arose to test a prototype natural language personal assistant system designed to control certain secondary and infotainment controls of a luxury car. An instrumented road-legal test car suitable for an on-road study had been developed by a UK technology research agency. A study was therefore designed collaboratively with the research agency. The research agency was responsible for recruitment, ethics, experiment set up, vehicle loan and logistics, insurance, risk assessment and safety.

It was decided to use video and remote links in Study B for reasons of space, and to make the 'Wizard' less obvious and intrusive. A laboratory was created to host the lead researcher controlling the 'script', a professional actor, and a second researcher assisting with information retrieval, with a third researcher (from the agency) travelling in the car. The system was designed so that the face view, gestures and speech of the participant, along with forward facing road conditions, were relayed by video camera to an office where a professional actor spoke the words of "the system" in an intelligent but fairly robotic way with a guideline script. This was because a pilot study on the first day (with six participants) had suggested that when the actor spoke in a totally natural humanlike way, drivers invariably perceived they were interacting with an *actual* human assistant thus reducing the relevance of the study. A 30-minute journey route with a wide variety of road conditions was planned as described in Section 5.5.3 below. This duration was chosen for pragmatic reasons, having proved difficult to recruit participants for longer than 90 minutes in total (the procedure for each participant was to consist of the drive plus induction time before plus two interviews, making 90 minutes total). The 'Wizard' had a basic script common to all participants, but instructions, voice or delivery could be changed at any point where it was advantageous to explore participants' reactions qualitatively (and safely). This is explained in the 'Method', Section 5.7. The 'Wizard of Oz' format permitted the flexibility to surprise drivers with occasional highly intelligent 'sentient', proactive or empathetic behaviour, as well as erroneous, confusing, frustrating behaviour. This was considered advantageous in maximising un/natural-feeling scenarios and associated data capture.

5.5.3 Driving routes

In study A, each recruited driver's journey was required to be at least one hour, to overcome any 'unnatural' behaviours and possible awkwardness. To make the journey conditions and interaction scenarios as naturalistic as possible, all the journeys in Study A were journeys that the drivers had already planned to undertake for their own personal, family or commuting purposes (not business). No set route was planned. In order to study a wide range of scenarios and make between-subject comparisons, all journeys were required to contain all the following elements: parking and exiting a purpose built car park, urban travel in a built-up area with pedestrians and some traffic congestion (waiting at traffic lights, junctions or behind other cars), peri-urban roads (40 to 60mph limited single or dual carriageway), and motorway driving ('freeway' in the US; with speed limits over 60mph). The 'destination' was any destination required by that driver. Four observations were done in the USA and four in the UK to increase exposure to possible cultural differences.

In study B, the route had to be planned because of the wireless and internet technology used for the observations (which required a strong mobile data signal), and for insurance and traffic reasons. It was therefore the same for all participants. A map was printed for the in-car researcher to give directions from. The route began with exiting a parking space in a purpose-built car park in Milton Keynes, travelling through two miles of urban built up areas with pedestrians and some traffic congestion, five miles of peri-urban roads with speed limits of 40mph and many roundabouts, one three-mile dual carriageway road with a speed limit of 60mph, and one short two-mile section of motorway with speed limit of 70mph. There were no stops made en-route and the driver returned to the same car park at the end of the test and parked in a suitable parking space. The tests took place in December between 10am and 7pm so around one third of the participants drove in darkness, the others all drove in daylight or semi-daylight. Again, this increased exposure to different conditions and interaction scenarios, and was considered advantageous. The quickest time it was possible to complete the route was 24 minutes. The longest time taken was 66 minutes because of school and commuting traffic.

5.6 Sampling and recruitment

Naturalistic ethnographic studies are not intended to produce generalizable results and typically use in depth observations in combination with small sample sizes (Cycil, 2016) usually fewer than 12. Individual characteristics are less important in exploratory qualitative research than other kinds of research (Patton, 1990). Objectivity in ethnography can however be maximised by ensuring participants are 'representative' (Dewalt and Dewalt,

2002) and by ensuring they are ‘culturally competent in the topic being studied’ (Bernard, 1994). Therefore a 50/50 balance of male and female participants was sought in both studies. Similarly, only experienced drivers were sought aged 25 to 75, who drove at least 50 miles per week and owned a car, to fulfil the criteria of ‘cultural competence’.

Sampling and recruitment Study A

Ten hours driving observation time was considered appropriate to obtain sufficient observations, based on the pilot study experience on the first day. Therefore, a sample of eight drivers was considered appropriate for Study A. Four observations were done in the USA and four in the UK to increase exposure to possible cultural differences. The drivers were recruited using a ‘snowball’ strategy combined with Maximum Variation Sampling (Patton, 1990) using a garage owner and a local neighbourhood website to contact participants by email. Drivers were asked if they were due to undertake any round trips of more than 60 minutes in the coming week, and if they would be content to have a researcher travelling with them. The driver was then telephoned to explain the research method and check what road conditions were anticipated. If the route planned met the criteria above, the driver was selected. Eight drivers were recruited in this way. No incentive was offered.

Sampling and recruitment Study B

A sample size equating to ten hours of observation time was also sought for Study B. Because the journeys were predicted to be shorter (around 30 minutes in normal traffic), 20 drivers were sought. They were not the same drivers used in Study A. They were recruited by a specialist external market research agency. Participants had to provide documentary evidence that they personally owned a premium or luxury brand car (the types of vehicle anticipated to feature voice assistants in the near future) and agree that they were the required age and drove the required weekly mileage. This was therefore a homogenous sample (Patton, 1990) as commonly seen in focus groups. An incentive of £50 was given. Most participants were from the city where the test route was conducted, meaning the roads were familiar to them. The test car was a 2009 luxury saloon.

5.7 Method

Study A: Drivers interacting with ordinary cars

A brief preamble was read out and the purpose of the research explained (the naturalness focus was not revealed until the end). The researcher observed from the passenger seat taking field notes in a note pad. One side of the paper was used for direct observations and sketches without interpretation. The other side was used to record more interpretative or reflective thoughts, as in DeMunck and Sobo (1998). The study was primarily observational. Questions were only asked if an interaction was unclear, had failed, or was suspected to be particularly natural or unnatural-feeling. In totally stationary situations like traffic queues or traffic lights, more general questions and clarifications were allowed if there were no concurrent driver-car interactions. Otherwise the researcher simply observed.

Study B: Drivers interacting with simulated future cars

The instrumented 'Wizard of Oz' in-car assistant was initialised using the various loudspeakers, video cameras, a Bluetooth mobile phone connection, and the internet remote meeting software WebEx. In the laboratory, three computer screens were arranged in a row showing (1) the live forward-facing road view (2) driver's face and upper body from the side (3) driver's face from the front. The participant was greeted at the car by the in-car researcher who explained the procedure and was asked to familiarise themselves with the car. Participants were told they would be using an advanced prototype built-in 'talking car' computerised assistant with access to diary, emails, traffic, GPS, local information etc. They were told they could interact with the system by voice or by any gesture they wished. Participants then drove the car for approximately 30 minutes on public roads on the predetermined circuit around the city of Milton Keynes. Navigation instructions were given by the in-car researcher. At the end, the driver was guided to a parking space.

The 'Wizard' could respond to very subtle cues (such as facial expressions from drivers, picking up on key words overheard in conversations inside the car, even mimicking the driver's regional accent) enabling relevant and helpful suggestions to be offered. The research team was able to access real time information on traffic, route, and points of interest and proactively tell the driver the price of petrol at passing petrol stations, or types of coffee at the upcoming coffee shop, for example. When traffic was encountered (which it was on every test), the assistant would offer to email someone about the delay. The style of speech was varied within each journey, from more slow and formal at the start, to a more conversational style near the end of each journey, to explore differing reactions. To explore a diversity of interactions, deliberate frustrating errors were made by the 'Wizard' (near the end of the journey so as not to break the illusion of optimal future technology for the most part). These were usually errors in comprehension, sometimes factual, sometimes mimicking a computer malfunction. The lead researcher took notes on a computer in the laboratory while the in-car researcher asked questions 'in the moment' about how the various system features felt to use, in the style of a 'Think Aloud' (Ericsson and Simon, 1984) when it was considered safe to do so. At the end the lead researcher conducted a short interview immediately after the participant exited the car, more explicitly asking about their *naturalness* perceptions of the system. After this were participants told about the nature of the 'Wizard of Oz' simulation.

FIGURE 5.1 A US DRIVER BEING OBSERVED WHILE DRIVING

5.8 DATA ANALYSIS

To control bias, the observational findings were discussed with the other two researchers each evening after the day's tests were complete. This took the form of a debrief presentation where each researcher presented their own observations and interpretations followed by reflective discussion to agree the most probable meanings and any naturalness implications. The observations were then analysed using Thematic Analysis (Braun and Clark, 1990) using the same procedure to that described in the first study. Observations, interpretations and inferences were coded into similar theme categories (Patton, 1990) using coloured highlighter pens. The independent code checker was a psychology researcher from Oxford Brookes University. He reviewed the notes for two participants in both Study A and Study B, with their associated inferences. Data was not analysed to any existing framework, they were simply observations suspected to be relevant to the research question, arranged into themes in a 'bottom-up' way. An example of raw data is shown in Figure 5.2.

5.9 RESULTS

Eight drivers (five female, three male) with mean age 35 years (SD=8) participated in Study A in January and February 2015. These journeys had a mean duration of 74 minutes. Twenty-two drivers (11 female, 11 male) mean age 36 years (SD=3) took part in Study B in December 2014. Those journeys had a mean duration of 35 minutes.

An inductive approach appeared the approach most faithful to the data - themes were closely linked to the *data* rather than the questions that elicited it. Themes were ignored if not exhibited by at least 30% of participants *per study* (i.e. a valid theme could result from being expressed by 31% of participants in Study A *or* 31% of participants in Study B. Setting a more strict minimum theme citation level per participant (e.g. *every* participant must express a theme in more than 30% of their verbal data by word count, or more than 30% of the time of observation) would have been too restrictive and would have resulted in no themes of interest. This is consistent with the flexible but consistent approach advocated by Braun and Clarke (1990). Also, in view of the lack of data directly using the semantic 'natural' (partly as a result of it being mainly observational data), a more latent/interpretative approach had to be taken rather than a simple semantic approach (Braun and Clarke, 1990). The findings of both studies are presented graphically and thematically below.

Example ethnographic field note Participant 2. [NAME]

Confusion re gear selector. *Gets comfortable, Seems confident.*

Nods to voice info. Looks straight ahead (no craning). Nods during info (*to herself?)* to indicate understanding.

"Yes I understand you" is her first comm. So *she must believe it is NLU.* Says sound is a bit muffled.

Nice bright "yes". Again she looks unclear whether she needs to acknowledge every statement.

Says 'yes' instinctively to her name. So is talked over by VID with the time delay. This happens several times.

Unsure how to acknowledge informational statements requiring no action (occasionally says "OK/Thank you")

Challenges system: "no, I would like to make a call" – *quite natural language is arising.*

Generally "yes please" is her response.

Critical if the system does not respond in a logical way.

Call ended – "thank you". Licks lips while thinking. Looks unsure at first.

Confidently/loudly starts calling "Steve" as a call sign for the car, unprompted.

Speaks deliberately but politely. "I would like you to set a reminder please".

[VID makes Deliberate mistake]: "No Steve". Corrects Bread for Fred. Laughs at error (*forgiving?).*

She seems quite forgiving – says to Derek that if she got a text reminder to buy a birthday present for Bread she would know that it meant fred. (*cf 'technology should behave as a likeable person'*)

"Steve?". Reverts to polite clear 'Thank you' after every response.

She initiates dialogue with planted Starbucks questions. "Yes please". *Unsure if she would naturally have asked a question of this type.*

Near end she is confident enough to address Steve and straight away ask for something, rather than waiting for the system to say 'yes?' e.g. "Steve can I…". Says thank you.

Impressions: Nice concept – this is her life, she has two lives, home and work, she may think about both interchangeably so system is useful. How describe to friend: Makes use of wasted time in car.

FIGURE 5.2 EXAMPLE ETHNOGRAPHIC FIELD NOTE (TYPED) FROM STUDY 5B

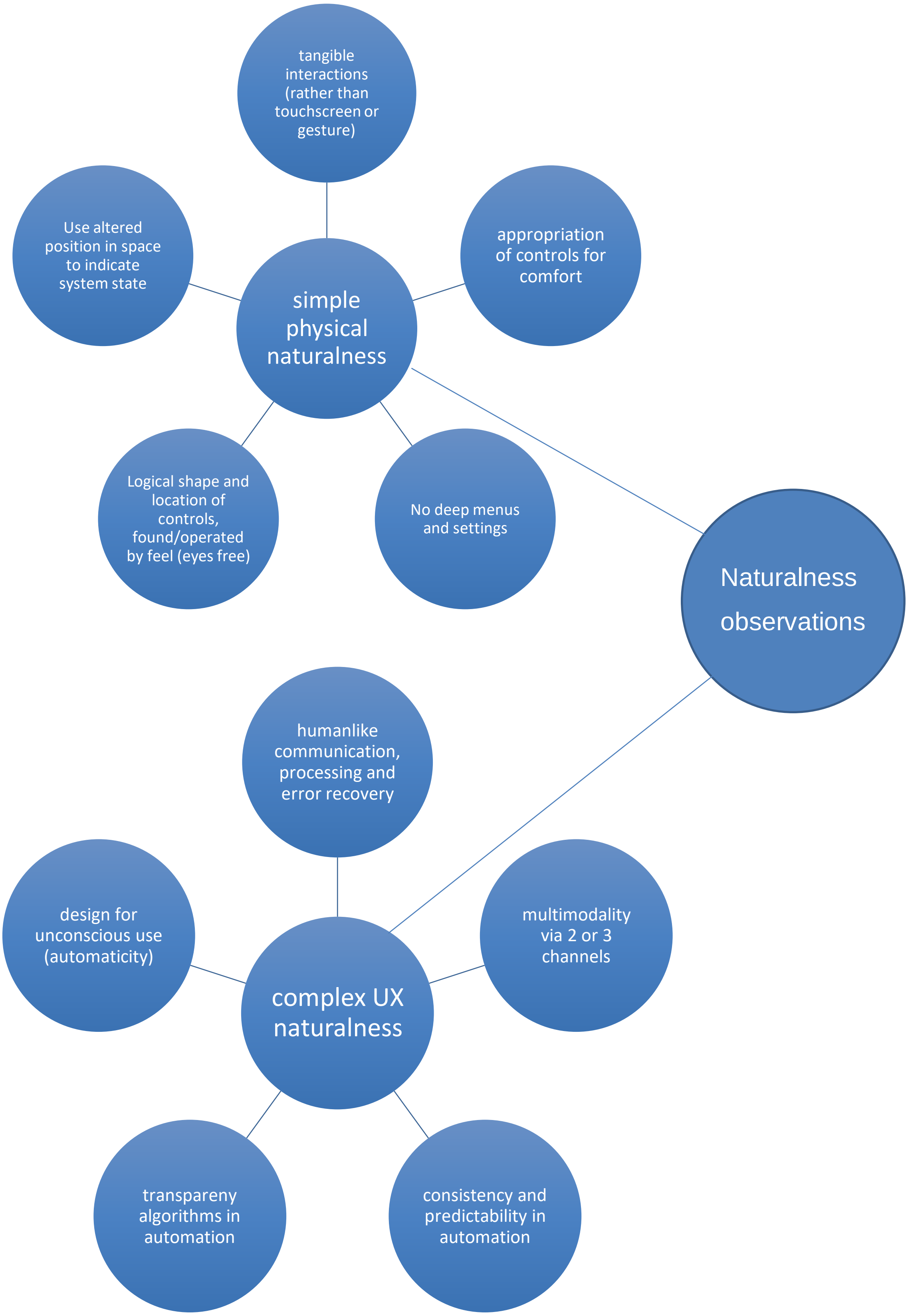


FIGURE 5.4 THEMATIC MAPS INDICATING SPREAD OF RAW DATA ACROSS TWO BASIC UNDERLYING THEMES OF PHYSICAL NATURALNESS AND COMPLEX NATURALNESS USER EXPERIENCE

The raw data suggested two main groups of themes – one group around ‘simple’ natural interaction relating to physical design and usage considerations, the second group around more complex user experience design considerations. Secondary interactions were clearly not the main objective of ‘real’ journeys; in Study A there was a rather limited amount of ‘naturally arising’ activity observed which could be classified as either ‘natural-feeling’ or ‘unnatural-feeling’ according to drivers’ facial expressions, body language, verbal expressions or answers to clarification questions. This amounted to no more than 20 recordable definitive ‘naturalness incidences’ across nearly ten hours of observation. Perhaps because drivers were so familiar with some of their secondary systems and their control locations, and appeared to be able to operate them semi-consciously and without looking, there was often simply no discernible facial expression or visual activity to observe. **However, this apparent limitation may paradoxically also be the most important result from this study – true natural interaction is possibly unobservable and not even consciously experienced by drivers**. After further analysis the naturalness incidences are summarised below thematically, but they were not sufficient to build a complete framework from. This is because it appeared that not every possible un/naturalness scenario had been *encountered* or *recognised* in the journeys driven and within the limited scenarios occurring on those journeys. This hesitation is because thematically, the naturalness observations corresponded to fewer than half of the ‘themes’ revealed by the previous two chapters’ studies. Therefore, it is possible that many of the opportunities *required* for the phenomenon of naturalness to arise, had simply not occurred.

The more ‘interventionist’ design of Study B encouraged and, arguably, *forced* verbal-auditory human machine interactions throughout each journey from the start, many being secondary control interactions. Estimating drivers’ intent and perceptions was therefore more straightforward. Tone of voice, wording and phrasing were easier to interpret than the more subtle eye/head/body movements and occasional facial expressions in Study A. Therefore, many more data relevant to the research questions were obtained, though again not sufficient to confidently propose a full framework for ‘every aspect of natural interaction with secondary controls’. This is because they were skewed towards a particular type of interaction i.e. hands-free verbal-auditory interactions with a highly intelligent conversational interface – as a result of the ‘speaking car’ experimental set up. Some naturally arising head gestures may have been missed because of difficulties in discriminating them from involuntary head movements caused by forces in the moving car. No participant appeared to perceive, when questioned just after the drive, that the intelligent ‘system’ had been a real human being not a computer. This arguably gives the study some face validity. The data may be summarised by the following 12 observations each derived from actual occurrences.

1. Physical factors were not observed to be fundamental in naturalness apart from distinctiveness of shape, logical location, and the desire for some physical controls rather than an entirely screen-based interface

Little evidence was found to explicitly support the importance of 'weighty' physical feel during on-road interactions, other than the naturalness preference for having some physical controls in the first place. *Control shape* however appeared important in naturalness and the many problems with touchscreen system usability appeared to demonstrate unnatural interaction (which may be as a result of *lack* of physical feel). This does not mean that feel is not important in naturalness, it just suggests it was *unobservable* – and as discussed above, this may even (paradoxically) indicate that 'feel' is *so* natural in automotive interactions that (for experienced drivers at least) it is an instinctive unconsciously experienced phenomenon. In other words, drivers are unable to 'observe' the importance of feel in their own interactions. By contrast, drivers appeared to find touchscreen operation quite demanding, both physically and cognitively (suggested by eye movements, glance behaviour and facial expressions) particularly when there were multiple settings and menus. At such moments their attention was seen to be markedly drawn away from the road. Drivers also appeared to have fairly low tolerance for touchscreen 'misunderstandings' and touchscreen 'insensitivity' (not registering a touch input) appearing to get angry and frustrated very quickly.

2. Appropriation of physical controls and artefacts for comfort appeared to be natural behaviour

Many drivers appeared to *appropriate* physical controls and artefacts in the car, for their own comfort. Examples included using the gear lever as an armrest or hand rest, using the handbrake as a hand rest or 'stress reliever', using door pulls or window sills as arm rests. This appeared to be 'instinctive' behaviour demonstrated by nearly every participant.

3. Hand-arm gestures did not appear instinctive or natural in the car

Although all participants were told that the intelligent system in Study B could respond to gestures as well as voice, there was only one instance of a hand/arm gesture used by a driver. Where a visible gesture was made by the driver (usually a head nod or shake as a confirmation or negative, together with verbalised 'yes' or 'no') the 'Wizard' *always* responded to acknowledge it and encourage further gesture. Nevertheless, no subsequent or expanded gestural communication was observed. The only occurrence of spontaneous hand/arm gestures was when drivers were communicating with *other drivers* on the road, *pedestrians*, or the *other people* in the car. This suggests that gestural interaction in the car is only natural with other humans or with another gesturing entity.

4. Change of controls' physical position in space may be a natural cue or 'feedback' for drivers to judge its state

Several instances were observed of indicators and wipers being left on unintentionally at motorway speeds. This situation appeared to be exacerbated by many recently manufactured column mounted levers now not changing *physical* position when activated. In older cars the physical 'out of resting position' of the indicator lever may have attracted the driver's attention even when the other two feedback modes (clicking sound and indicator lamp/wiper sweep) were imperceptible because of automotive noise, glare or competition over the visual mode of feedback This also suggests multiple redundancy and multimodal feedback may be natural in the automobile (See finding 7 below).

5. Automaticity in secondary control operation appeared natural

There appeared to be multiple routine 'automatic' or semiconscious driver interactions with frequently used tangible secondary controls (hard buttons, dials, switches and stalks) and this appeared to be a 'natural' or instinctive phenomenon. Often no noticeable change in facial expression or head posture was noticed during this interaction, and the driver was able to sustain conversation (unlike with screen-based interfaces where conversation tended to stop). When questioned about a recent input action of this type, the driver was not always aware of what action they had just done. Automaticity has often been viewed negatively in the context of driving, but this finding suggests the opposite hypothesis deserves investigation too – that secondary interactions which are familiar, eyes-free and intuitive, help drivers' attentional/cognitive resources to remain on the primary driving task, rather than being diverted to the secondary activity.

6. Controls located by 'hand' with low visual demands appear natural

Drivers' hands often appeared to instinctively 'hunt' or 'feel' for secondary controls, not looking directly at them, but using shape and location as a guide. The hand would explore by touch, seeking a familiar shape or texture in an approximate target area. This appeared to be a 'natural-feeling' activity. Where a dashboard had two similar shaped buttons in the same location, mistakes were sometimes made in activating the wrong one (for example a round menu navigator/selector knob was confused with a nearby round fan speed knob on a Chrysler car). Drivers did not appear to know where all their controls were located, especially rarely used controls, and this could also cause distraction when the driver needed to look over the whole of the dashboard to identify a particular control. This suggests that a wholly physical dashboard would not always feel natural in a car with many features, but that some of the most frequently used controls and settings should remain physical rather than migrating wholly to the screen-based interface.

7. Feedback in two or three modalities (redundancy) appears to be natural

Controls which demonstrated two or more modes of feedback appeared to be perceived as more natural. Most commonly this was a ‘visual indication of state’ combined with an audible click or beep. The ‘visual indication of state’ might be an LED or graphic, but it can also be a simple change of control position in space (as in (4) above). Directional indicators provided a good example of tri-modal feedback.

8. Semi-automated or occasionally automated controls appeared unnatural

Semi-automated systems (i.e. those which were *occasionally* automated, or only displayed automation over *part* of their function range), appeared unnatural. Although not strictly a secondary control, electronic semi-automated handbrakes provided a consistent example. Drivers often appeared not to be aware what state the electronic handbrake was in (whether on, off, automatic, manual, engaged, released). Lacking the familiar visual cue of a large lever being either ‘up’ or ‘down’ in space, allowed less natural judgment of state. Drivers appeared especially confused by the lack of *consistency* in handbrake automation (such devices sometimes activating automatically in some scenarios but not in others). In several cases this created lack of trust in the control, and compensatory measures, or even ‘disuse’ (using the transmission ‘park’ mode instead, even leaving the engine running in one case).

9. Transparent humanlike algorithms appeared natural in automation

Where automated systems exist, drivers appeared to find them more natural if they could understand the algorithm behind it. For example, an automated night-time high beam assist function was forgiven for occasional overcautious deactivation (when passing a bright street lamp for example) because that driver could understand the underlying algorithm (in this case a simple forward-facing light sensor at the base of the windscreen). Transparent algorithms appeared to contribute to understanding of complex systems, and their naturalness, perhaps by building mental models, trust or metaphors. At the current stage of technology, simple algorithms and metaphors appeared natural (compare this ‘simple’ high beam assist algorithm with the electronic handbrake algorithm above, which was complex, inconsistent and conditional, and did not give the driver ‘natural’ understanding or control)

10. Hands-free operation via voice appears to feel natural if well executed

Drivers appeared to find the low visual and low physical demands of the intelligent voice assistant very natural within an automobile, if paired with near perfect natural language recognition. Drivers appeared to operate their cars while juggling multiple life tasks (such as talking, telephoning, or looking for something in the cabin) all of which tend to compete over the visual and physical (manual) modalities. Observations suggested that within two or three turns of ‘successful’ conversation with an intelligent automobile, drivers felt comfortable and natural about this way of interacting. Perhaps one unnatural aspect was the lack of eye contact – drivers’ eyes were observed to look in various unpredictable locations while talking (at the rear-view mirror, at the cluster, even out the side window). Even when the *head* was oriented to the road, the *eyes* were sometimes observed to dart side-to-side and up-and-down as if ‘hunting’ for eye contact or affirmation. Considering that speech control was developed to keep drivers’ eyes on the road, this is a surprising failure in interaction design.

11. Humanlike tone of voice and communication styles appeared natural

Humanlike tone of voice, delivery and politeness all appeared highly important in drivers’ perceptions of an intelligent car system as natural. Smiles and laughter were frequently observed when the car reacted in very humanlike ways (such as when the car appeared to ‘learn’ from its driver, combine two pieces of information to make a valid inference or helpful suggestion, or mirroring of communication mannerisms). ‘Self-learning’ (the ability of the car system to proactively learn the driver’s preferences, intent or routines) seemed to be perceived as an especially natural aspect. Within two or three turns of successful dialogue at the start of the driver’s interaction with the Wizard of Oz voice system, drivers were observed to use apparently very ‘natural’ informal and nuanced speech (despite the problems a normal speech engine would have in recognising such speech forms). Examples were the driver saying ‘No, you’re fine thanks’ and ‘Not right now thanks’ as polite euphemisms for ‘No’. Expressions of ‘yes’, incidentally, were much less varied – restricted to ‘Yes’, ‘OK’ and occasionally ‘All right’. Successful task completion led to drivers attempting more complex speech interactions for example combining two requests into a single utterance, or executing a search *and* filter request rather than doing these actions separately. An even more complex form of utterance was observed on three occasions whereby the driver would ask for one or two pieces of information *related* to a goal (for example the distance to a place, or identifying a place of interest), before making a decision and then issuing a command which implied *contextual* knowledge of the recent dialogue using referential pronouns such as ‘can you direct me *there’* or ‘can you call *them*’). Drivers were observed to be pleased when such interactions were successful.

12. Humanlike recovery from error appeared natural

Mistakes appeared to be quickly forgiven if recovery was logical and humanlike, especially where the reason for the mistake was obvious to the driver. If errors were handled in a *non-*humanlike way (for example using computer type error messages, or the nonsensical error messages simulated in Study B) drivers appeared to lose patience very quickly and would often attempt to end the session by any means possible.

5.10 Naturalness framework which emerged from the ethnographic observations

The findings obtained were not considered sufficient to be described as a 'complete' framework, as explained above, because there appeared to be insufficient findings to corroborate *every* theme from previous frameworks. However, the findings were considered to be essential for the triangulation process explained in Chapter 6. Following thematic analysis, they were summarised below as a list of naturalness 'design considerations' grouped as 'physical', 'usability', 'automation' or 'humanlike communication' considerations.

Physical considerations

- Distinctive shape, texture and location of controls may contribute to natural feel;
- Certain physical controls may perform secondary roles as hand or arm supports;
- Hand-arm gestures are probably not a natural way to control an automobile;
- The physical position of a secondary control in space may be a natural feedback device if its activated state is a visibly different position to its 'off' state.

Usability considerations

- Regularly used secondary controls should be designed for semi-conscious operation;
- Controls should be usable without the driver needing to look at them because eyes-free operation of controls appears natural;
- Feedback in two or three different modalities (i.e. sensory channels) may be natural, especially where it compensates for feedback channels impaired by noise and glare.

Automation considerations

- Natural-feeling automation may be permanent and operate competently across the system's full capacity; automation should not be occasional or partial; if this is not technically possible it may be more natural not to have any automation;
- Automation algorithms need not be complex to feel natural – simple humanlike, transparent and predictable algorithms may be preferable.

Humanlike communication considerations

- Hands-free voice operated systems have the potential to feel natural in the car but only with near perfect natural language recognition and understanding of nuances;
- Humanlike qualities of politeness, empathy, learning and proactivity may be perceived as natural in intelligent secondary systems;
- Humanlike recovery from error may be perceived as natural in intelligent secondary systems.

5.11 DISCUSSION

Ethnographic participant observation of ordinary drivers driving cars was necessary to permit triangulation (Decrop, 1999), to ensure the research was credibly human-centred (Giacomin, 2012b), because it has been suggested in the literature that naturalness is a situated and occasioned phenomenon and therefore cannot be fully understood by asking people about it or conducting laboratory studies. Furthermore, some of the subtleties, gaps and contradictions in the themes obtained from the first two studies needed clarification. In particular the themes concerning naturalness in social, humanlike and intelligent driver-car interactions needed to be better understood. In part, this appears to have been achieved.

Before detailed discussion of findings, consideration must be given to the way dependent variables were measured and interpreted, which was somewhat different to the other studies. Firstly, despite the advantages of ecological validity and naturalistic context, the observer/researcher's role was necessarily rather more subjective than in the other studies. This is because, in most cases, the un/naturalness of any interaction had to be first judged by the researcher, and then explored by the researcher using rather leading questions. Interpretations of facial gestures and body language varies between individuals, just as drivers' success or failure in meeting minor goals cannot always be judged by an observer. Because the study was observational (i.e. primarily visual) and interactions were generally silent and private, interpretations of naturalness were likely have been skewed towards 'visible' manual and bodily interactions (i.e. potentially ignoring the social or higher processing elements of interactions). Similarly, some 'skilled, learned' natural interaction is likely to have been missed, because it is hard to judge if an interaction is skilled/learned or just 'easy'. Finally, naturalness considerations about drivers' mental models and metaphors cannot possibly all have been captured because no observation can read the minds of participants. Suchman's (2007) central theme was that machine operators do not set out with a fixed plan at the outset, but adapt and formulate it depending on the unfolding situation at hand. Observations suggest this might be equally applicable to drivers and cars.

That said, while there were insufficient observations to corroborate *every* theme from Chapters 3 and 4, the considerations listed above did largely accord with the previous two frameworks, and will facilitate their amendment during the triangulation process described in the following chapter. The study has also introduced some *new* possible naturalness contributors – control shape, semi-conscious operation, multiple redundancy, and appropriation. Appropriation is a common concept in sociology and product design generally but rarely discussed in the automotive domain. The study has especially helped to clarify the subtleties of the ‘socio-intelligent’ constructs in the two draft frameworks – where it had been difficult to demonstrate coherence *within* themes and distinctiveness *between* them (Patton, 1990). The data suggests that a natural-feeling intelligent car system should be polite, subservient and fairly (but not completely) humanlike. It should learn from the driver by monitoring, and use this to exhibit empathetic ‘mirroring’. Mirroring may relate to the driver’s habits, preferences, or speech style.

Some elements of *physical* interaction appeared natural in driver-car interaction (despite the likely physical/manual bias discussed above). Drivers often continually adjusted round dials such as temperature and volume throughout a journey. This may be related to the satisfaction of ‘full control’ as featured in the two frameworks. However, the primary importance of physicality, weight and ‘springiness’ suggested by previous studies, was not demonstrated here. Controls where there was a clear ‘affordance’ (Norman, 2013) or visual ‘positional feedback’ (i.e. the passive position of the control in space giving feedback as to what mode it is in) appeared more natural. In particular the *shape* and *location* of secondary controls appeared important in drivers’ choosing the correct control, thereby meeting their need or goal, and affecting their judgment as to whether the interaction felt natural or not. This is especially important as so many secondary controls appeared to be naturally operated ‘semiconsciously’ or ‘blindly’ with the driver keeping their eyes and attention on the road, or on other passengers in the car. Banks of identically sized and shaped controls close together would seem unnatural in this regard, yet these are a feature of many current car dashboards. More problematically, manufacturers are currently moving more secondary functions (such as ventilation) completely onto central infotainment screens (whether graphical or menu based) which violates this naturalness preference for distinctive control shape and position. Hand-air gestures do not appear to be instinctive when interacting with the car. At most, head movements to communicate ‘agreement’ or ‘disagreement’ might feel natural in a car. Therefore, it cannot be assumed that an entirely screen based interface can be made to feel more natural by adding gesture control.

The naturalness frameworks’ themes of ‘usability’, ‘low visual demand’ and ‘low cognitive demand’ were apparently affirmed in part. Included within ‘low cognitive demand’ might be a presumption of not providing too many functions or settings, because they appear unlikely to be used or understood in practice. Cognitively demanding multimodal screen-based systems appeared to have high visual and cognitive demands and often broke the semiconscious ‘control loop’ by momentarily fully diverting attention off the road. Furthermore, mode confusion was observed in multimodal screen-based systems. Mode confusion is considered to be a contributory cause of many aviation and automation supervision accidents (e.g. Sarter and Woods, 1995).

When the intelligent voice system exhibited natural language outputs, natural language recognition, appropriate responses and sensible error recovery, the overall interaction style quickly became a productive natural language dialogue and drivers perceived it as very natural. It is possible that the current reported low uptake of voice systems in cars is simply because of poor first experiences and system design. If a driver is not understood, ignored or confused in the first few interactions, they may never engage with that system again. However, the evidence here is that car voice systems which *do* listen, understand and process intelligently and attentively can feel highly natural in a very short time. This is hardly surprising given humans have evolved to communicate mainly in the verbal-auditory domain (or at least, face-to-face). Much as in Reeves and Nass (1996) it appears that drivers readily stereotype verbal-auditory technology based on its voice, and in some respects 'naturally' treat an intelligent car like another human (rather than a machine). Examples of this include drivers referring to the system as 'she' or 'he' rather than 'it', body language or facial expressions indicating human-human conversation (like lip pursing, attempted eye contact – but not gesture), use of colloquialisms and politeness (such as "not right now thank you") and use of humour. Designing a fully humanlike voice system at the present time may however create false impressions of fully human intelligence (Shedroff and Noessel, 2012) and in any case, may not be technically achievable at an affordable cost for many years.

Naturalness did not appear to be wholly correlated with intelligence or capability, only with perceived *competence* and *consistency*. Consistency in engineering appeared to be important in naturalness. The data suggested that if an intelligent secondary control system can control the headlights then it should also be able to control the wipers and windows. If a computer controller controls one system intelligently, it should demonstrate that degree of intelligence in every other system it controls.

Comparing these findings to definitions of natural interaction from the literature, a similar pattern is found to that in the existing studies' discussions. The finding that naturalness may be associated with control shape and position may relate to the naturalness of tangible and embodied interaction (e.g. Hornecker, 2011) and possibly also the notion of affordance (e.g. Norman, 2013) which he and others describe in what may be thought of as 'natural' terms. This corroboration is strengthened by the observation that *digitally represented* interactions within the car (e.g. with touchscreen or WIMP interfaces) appeared to feel unnatural to many drivers. As in the previous two studies, naturalness appears to be enhanced by mimicking certain human-human communication tendencies such as speaking/listening, mirroring, colloquialisms and context referencing – and perhaps might be further enhanced by somehow mimicking eye contact. However, non-contact gesture on its own did not appear to be natural, so it would be inappropriate to assume all aspects of human-human communication can be mimicked naturally.

Most interestingly, the apparent automaticity and unconscious execution of natural interactions may theoretically be a logical extension of the Goodrich and Olsen (2003) interpretation of natural interaction: if natural interaction uses mental models that are *so* well calibrated, stimuli that are *so* familiar, and response generation that is *so* well-practised, perhaps the whole interaction is processed unconsciously and 'automatically' almost like a natural 'instinct' (Malizia and Bellucci, 2012). In this regard, the Wigdor and Wixon (2011) standpoint would still appear to apply because such interactions would likely only feel natural or instinctive to the experienced driver, not the novice. Thus 'unconscious' naturalness cannot be a property of the interface itself.

5.12 CONCLUSIONS

Eight participants were ethnographically observed operating secondary controls in current day cars, and 22 participants were observed operating various secondary functions in a simulated futuristic car controlled mainly by voice, both on public roads. Observations were made relevant to the research question regarding what appears to feel natural and unnatural when interacting with a car's secondary controls. Because of the combination of current day and futuristic prototype secondary system controls, a variety of possible scenarios was observed with reasonably high levels of ecological validity. The two previous naturalness frameworks have been partly corroborated by these findings, despite the limited data obtained from Study A. Some clarifications and logical distinctions may now be added to the themes of the previous frameworks during *triangulation* in the next chapter. Some additional naturalness design considerations have been suggested but they are believed to be an 'incomplete' framework even after 23 hours of participant observation. This was probably because not every naturalness scenario could arise 'naturally' in naturalistic driving situations where secondary interactions arose mostly by chance or by the free will of the driver. The main contribution of this study relates to the possible naturalness of unconscious control, control shape, appropriate location and physical contact; the naturalness of designing for 'blind' and 'semi-conscious' interactions with feedback in more than two different modes; the naturalness of 'positional feedback' as an indicator of system state; the naturalness of competent consistent automation with a clear metaphor/algorithm; and the naturalness of humanlike communication stereotypes.

CHAPTER 6

TRIANGULATION OF THE THREE FRAMEWORKS

6.1 AIM

Two complete draft frameworks for natural-feeling driver-car interaction with secondary controls have been proposed from the Contextual Inquiry and the exploratory design workshop studies, along with a ‘partial framework’ in the form of a list of naturalness ‘design guidelines’ from the ethnographic studies. There were some common higher order themes across all these frameworks concerning feelings of control, physical connection, dashboard layout and usability; and perceptions of humanlike sentience and humanlike communication. The wording of the themes differed slightly between frameworks, and some themes only appeared in two of the three frameworks. This is most likely because they were derived independently from different datasets with small to medium sample sizes, and because each study had its own experimental biases, rather than because the measured construct was different in each study or the samples were not representative of the intended population.

Specifically, it was uncertain whether ‘comfort’ was an essential contributor to naturalness or a more general human need, and whether ‘familiarity’ and ‘predictability’ were central to naturalness or were more general expectations about *any* human machine interaction. Physical feel (such as ‘weightiness’, and ‘tightness’) had been observed to be important in two studies, but not in the third. Usability issues appeared rather more prominent in the first two studies than in the ethnographic studies.

A unified thematic framework was required, to use in quantitative rating scale development. A method was therefore sought which could achieve consensus across three sets of themes obtained by rather different qualitative methods, so that themes could be confidently converted into questionnaire items for the final study described in Chapter 7.

6.2 RESEARCH METHOD

Triangulation is a common strategy in mixed methods qualitative research where at least three studies have taken place using the same research question. Based on the topographical metaphor, triangulation implies that a single finding is considered from three different independent viewpoints or sources. Triangulation can show that several independent sources converge on the same findings, or at least, do not produce opposing findings (Decrop, 1999). The four basic types of triangulation are *data triangulation, method* *triangulation, investigator triangulation* and *theoretical triangulation* (Decrop, 1999). Triangulation can strengthen a study by combining methods (Patton, 1990). It is considered to add 'trustworthiness' and 'believability' (Golafshani, 2003) to qualitative research. Triangulation is considered by many qualitative research methodologists to be the equivalent of 'rigour' in quantitative research (Decrop, 1999). From Henderson (1991):

> *"[Triangulation] guards against the accusation that a study's findings are simply the artefact of a single method, a single data source, or a single investigator's bias" [p11]*

An integral part of the triangulation process is to reflect on what *biases* each of the methods may have encouraged, and their possible effects on the results, in a process known as 'reflexivity' (Mruck and Breuer, 2003).

6.3 Key Methodological Decision

This research was considered most suitable for *method triangulation* (see Figure 6.1), and this had indeed been the intention from the outset, because data from three *different* methods were obtained with respect to the same research question, each dataset had been analysed independently, and each would have its own distinctive potential sources of bias. Based on a review of different triangulation studies, a matrix approach was chosen for the reflexivity, which considered experimental conditions and their potential effects under the headings of *physical conditions, elicitation techniques, social context,* and *temporal context*. This framework was based on the *PEST* analysis technique for business and market analysis (see Drummond and Ensor, 2006) and its many variations.

6.4 Objectives

1. To 'reflexively' consider potential method-induced biases by comparing each study's methods with any possible corresponding anomalies and skews in its findings;
2. To use this to corroborate, illuminate or discount each of the findings across the three sets of results in turn (Patton, 1990);
3. To discount any anomalous themes likely to have arisen from method-induced bias;
4. To create a unified framework from the three different frameworks using a logical matrix approach.

6.5 METHOD

Firstly, an analysis was performed of all the different study conditions and the role of the researcher (and intrusion effects) across all three studies. Any possible biases that these circumstances might have created were then ‘brainstormed’ with two non-automotive researchers from HCDI at Brunel University. They were chosen for their wider qualitative research experience and their ability to objectively ‘detach’ from automotive stereotypes.

The experimental conditions were considered under the headings of *physical conditions* (for example issues of place, artefacts, lighting, environmental conditions and physical stimuli), *elicitation techniques* (whether interview, discussion, artefact centred or passively observed), *social context* (whether solo or group activity, and the role of the researcher in the social setting) and *temporal context* (this being issues of timing and the balance of present day versus future and past considerations). These headings were considered to adequately represent all the key differences in conditions between the three studies. Biases could then be identified along with their corresponding potential effects on the data. This usually took the form of estimating which aspects of the findings were likely to have been *skewed*, or which findings were likely to have been underrepresented or overrepresented, using brainstorming and discussion techniques). This ultimately permitted various decisions regarding which findings to ignore, amend or retain during the subsequent data review. These decisions are summarised in Figure 6.2 and Table 6.1 below.

All the findings from all the studies were then reviewed with these considerations in mind, in a second matrix (summarised in Table 6.2). It was possible, in most cases, to track each naturalness theme from one study to the next, despite them being derived independently, because of the relatively subtle differences in wording between them.

The third and fourth columns of Table 6.2 were initially blank to allow the summary ‘triangulated’ theme wording and its rationale to be added. This was done after comparing each theme with the list of likely biases in Table 6.1 and discussing the implications and wording options with an independent researcher chosen for his strategic abilities and concise writing skills. The result was a triangulated framework of ten naturalness themes. This is summarised in Figure 6.3.

FIGURE 6.1 MODEL SUMMARISING THE METHOD TRIANGULATION

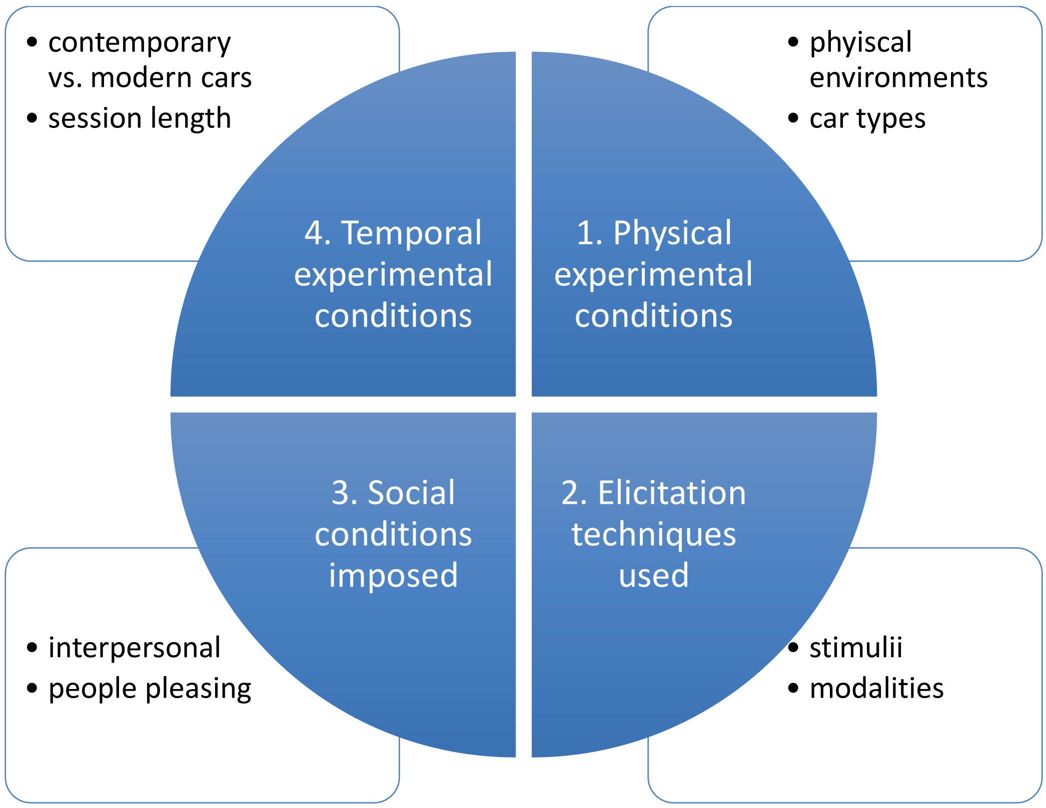


FIGURE 6.2 CONSIDERATION OF ALL THE EXPERIMENTAL CONDITIONS POSSIBLY LEADING TO BIASES

6.6 Results

As explained above, the most concise summary of the results of the method triangulation is given by the following two summary tables.

Table 6.1 Summary triangulation matrix illustrating possible sources of bias and implications

Aspect	**Study 1: contextual inquiry**	**Study 2: exploratory design workshops**	**Study 3: Observational ethnography**	**Possible biases and implications for interpreting and triangulating results**
1. Physical experimental conditions				
Physical environment of the study	Parked car in car park, real world daylight or night-time scenery out of windows. Variable weather. Wide range of cars; drivers' own.	Windowless artificially lit laboratory with multiple automotive stimuli (parked car and multiple dashboard components). Constant temperature.	Variety of road conditions, weather, participants and countries (UK and USA). For most, the roads were very familiar, in 3A the car was very familiar (the driver's own).	Outward visibility issues likely to be more salient in Study 1 (interviewed looking out of windscreen). Bias towards finding controls in dark in Study 2. Cabin temperature and climate also likely to be emphasised in Study 1. Study 3 most ecologically valid. Perceptions in Study 2 less ecologically valid.
Familiarity of car environment	Very familiar	Unfamiliar	Combination of very familiar (3A) and unfamiliar (3B)	Study 2 more likely to elicit usability issues: drivers know their own car's controls well but face multiple usability/learning issues in a new car/with new controls. Over-familiarity with own cars (able to operate semi consciously) may prevent insights and problems being raised in Studies 1 and 3A.
2. Elicitation techniques used				
Elicitation channel/modality	Structured interview questions + spoken (narrative) responses	Mix of interview, discussion and practical play/creation + spoken responses and physical creations.	Observations of physical activity only, occasional verbalisations Very few questions. The most ecologically valid study.	Study 1 may have 'left brain' narrative bias. Study 2 likely to elicit broader perceptions, less linguistic/narrative biased. Study 3 has potential to reveal the most 'natural' behaviour, but it may not always be interpreted as such.

Aspect	Study 1: contextual inquiry	Study 2: exploratory design workshops	Study 3: Observational ethnography	Possible biases and implications for interpreting and triangulating results
Stimuli and scenarios given to elicit responses	Dashboard visible throughout. Questions concerned actual past experiences and 5 driving scenarios.	Multiple multisensory stimuli (loose car controls, non-car controls, materials samples, a real car, multiple creative exercises.	No artificial stimuli or scenarios given in 3A. Both 'real'. 3B by contrast had numerous stimuli and scenarios given by the 'Wizard' script.	Study 2 likely to give deeper perceptually 'instinctive' least cognitively biased answers. However it may be biased towards physical feel and usability because of physical and unfamiliar stimuli. Study 1 biased towards actual past experiences and drivers' own cars. 3A most ecologically valid. 3B data should be used for 'socio-intelligent' aspects of the frameworks only (ignore other data – prototype)
Ability to probe in the moment, unpack interesting or unexpected responses (verbal protocols, probing)	Not possible (structured interview format designed to be comparable between subjects).	Very possible and done whenever opportunity noticed.	Not done except where an interaction failed or was causing visible difficulty due to safety considerations	Study 2 likely to elicit more instinctive responses, capturing contradictions and perceptions. The breaching exercise results express 'the unwritten rules' which may otherwise go unreported.
Use of the word "natural" and its derivatives in questions	Avoided in most questions to avoid biasing answers. Dictionary synonyms used.	Used in all sessions apart from first session	Used occasionally during observations, used frequently during debrief interviews	While studies 1 and 3 may have confused participants as to their purpose, less likely to have encouraged socially mediated 'people pleasing' answers.
3.Social conditions imposed				
Negotiation, sharing and cooperation; interpersonal dynamics.	No group dynamics – solo interviews between in-car interviewer and subject.	About 75% of the session required negotiation and co-creation, 25% individual	Study 3A had no sharing or cooperation Study 3B had cooperation with the 'confederate' travelling in car.	Study 2 biased towards issues of delegation and sharing? Study 2 dominated by louder, higher status subjects? Researcher intrusion in all. Less opportunity for people pleasing bias in Study 2 & 3.
Possible Self-selecting 'car enthusiast' bias	Likely - no payment incentive	Possible but wording and ad placement designed to appeal to ordinary drivers. Small voucher incentive	3A deliberately sought non-enthusiasts. 3B offered a market rate incentive : bias towards enthusiasts was reduced? (but not eliminated)	Possible bias towards car enthusiasts in all. Enthusiast and extreme user type opinion should be interpreted carefully. Ignore codes not shared by 30-40% subjects.

Aspect	Study 1: contextual inquiry	Study 2: exploratory design workshops	Study 3: Observational ethnography	Possible biases and implications for interpreting and triangulating results
4. Temporal experimental conditions				
Balance between contemporary car and future car (or intelligent car)	A balance of about 50% contemporary and 50% future by question type.	About 80% contemporary cf 20% future according to session plan.	Around 30% contemp vs 70% intelligent by number of participants in 3A vs 3B.	Study 2 likely to be biased towards current car paradigms (and physical controls?), Study 1 well balanced. 3b very biased to future and only one possible interpretation of 'intelligent cars' was considered.
Overall length of sessions	Average 37 minutes	Average 160 minutes	3A: Average 70 minutes but observation was less intrusive on 3A (normal journeys) 3B: 30 minutes plus pre- and post-interviews.	Study 2 likely to get wider range of beliefs and perceptions. Fatigue may be an issue at end of Study 2 leading to 'researcher pleasing' or 'uncontroversial' results. Results from second half of Study 3 sessions may be more valid after initial 'awkward unnatural' behaviour overcome.

While most themes and subthemes were retained at this stage, the whole of the raw data was reviewed using the list of potential biases from the right-hand column, to look for the type of biases predicted. Where a prediction had been made that certain studies might have over- or under-represented certain issues (for example studies where unfamiliar car controls were being tested) the study was reviewed with biases noted on a sheet of paper in a bold typeface.

Annotations were made in pencil directly onto the various thematic models which were then reviewed together in the next stage. This next stage was comparing each theme in detail from one study to the next and deciding whether to retain, amend or delete it.

Because this was a panel discussion taking nearly two hours with debate and joint decision taking, the whole process is not easily described on paper, however the key decisions are summarised in Table 6.2 below. The lead researcher had the final decision because of his greater familiarity with the data. This second matrix table compares the themes across all three studies showing their 'evolution' and any critical wording differences between studies.

TABLE 6.2 EVOLUTION AND COMPARISON OF THE THREE NATURALNESS FRAMEWORKS' THEMES AND THE CHOSEN 'TRIANGULATED' THEME WITH MAIN RATIONALE

Theme from Study 1 (Ch 3)	**Equivalent Theme From Study 2 (Ch 4)**	**Equivalent guideline from Study 3A/3B (Ch 5)**	**Final Triangulated Theme wording**	**Main Reasoning for Decisions**
A: Themes connected with familiarity and control				
[Expectations Of Functionality Reliability and Familiarity]?	Familiarity And Predictability	Expectation of familiar location/shape	Familiarity And Predictability (Theme 1)	Should be included in the framework because it appears to be essential for any driver-car interaction, particularly natural-feeling interaction. Should be shown differently to other themes.
Full Control And Manoeuvrability	Driver In Full And Ultimate Control	Control use may be semi-conscious but this may be seen as well-practised full control	Driver In Full And Ultimate Control [centre of the framework] (Theme 2)	Manoeuvrability is related to primary driving not secondary controls. "Ultimate" added to allow for automation/semi-conscious control. Control appeared central to driving, not just naturalness.
B: Physical Themes				
Direct Connection	Real-World Feedback or communication with reality	Possibly, multimodal feedback and feedback in three modes	Communication With Reality (Theme 3)	'Direct connection' implied mechanical linkage which was unintended. Replaced with 'communication with reality'. It is the 'reality' that appeared important (whether road or vehicle), not the mechanics.
Rich Skilled Physicality	Weighty Physical Sensations	Possibly: Air gesture not natural, position in space as feedback, controls support hand, touchscreen unnaturalness	Weighty Physical Sensations (Theme 4)	Wording simplified. 'Weightiness' appeared the predominant common factor. Not possible to prove that interaction was 'skilled' or even 'learned'. Implies the unnaturalness of touchscreens with no detents or physical/haptic feedback.
Comfort	Cabin Comfort and Sanctuary	Controls may be appropriated for comfort as hand/arm supports	(Deleted)	Deleted because the ethnography (and data in Study 2) suggested this was a basic human need unrelated to naturalness. Some drivers suggested natural-feeling cars should not be too comfortable. Poor face validity.

Theme from Study 1 (Ch 3)	Equivalent Theme From Study 2 (Ch 4)	Equivalent guideline from Study 3A/3B (Ch 5)	Final Triangulated Theme wording	Main Reasoning for Decisions
C: Usability Themes				
Vehicular Usability	Uncluttered Cabin Architecture	Controls should be located without looking	Uncluttered Cabin Architecture (Theme 5)	Expanded from the single construct in Framework 1. So many expressions of usability were recorded it was necessary to divide them into 3 logically separate components. All appeared essential and all were observed to some extent in the ethnography thus improving likely validity.
	Low Visual Demand	Eyes-free; Distinctive shape' multimodal feedback etc.	Low Visual Demand (Theme 6)	
	Low Cognitive Demand	Low mental demand	Low Cognitive Demand (Theme 7)	
D: Humanlike and (Future) Intelligence Themes				
Acts Like A Technical Co-Pilot AND	Humanlike Driver-Car Partnership	Transparent algorithms natural (?)	Humanlike Driver-Car Partnership (Theme 8)	Humanlike partnership was the predominant theme. Co-pilot suggested a 'second driver' - unintended. Observations from 3B strongly reinforced the theme of 'humanlike' intelligence being natural, and sense of perceived partnership when intel. is natural.
Humanlike Proactive Assistance		Consistency of automation (not partial/occasional); theme suggested by multiple tone of voice obs in 3B		
Intelligent Sensing and Understanding	Humanlike Sentience and Learning	Humanlike algorithms; politeness; empathy; self-learning	Humanlike Sentience and Learning (Theme 9)	'Sentience' better expressed the naturalness data than 'understanding'. 'Sensing' appeared too computer-like. Evidence from ethnography strong. Drivers found it natural when the car learned 'sentiently' from its driver.
Acts As Single Intelligent Being	[not present]	Consistency of intelligence?	[Deleted]	Not enough evidence from drivers/cars tested. Wording also misunderstood by some drivers during face validity test.. Meaning better conveyed by other constructs mentioning 'humanlike'/ 'sentience'. Poor face validity.
Vocal Information Exchange	Humanlike Verbal-Auditory Communication	voice dialogue with intelligent, listening, understanding, proactive car can feel highly natural esp if error handling and mirroring are humanlike.	Humanlike Verbal-Auditory Communication (Theme 10)	Strong evidence that humanlike verbal-auditory communication is natural way of communicating in the car if NLU is near perfect and errors are recovered from in a humanlike way.

6.7 DISCUSSION

The commonality of the themes across studies, and the predictions about possible biases and likely data skew, made the triangulation relatively straightforward. In two cases, poor face validity and inconsistent data led to deletion of themes from the final model, in tandem with bias predictions borne out by apparent skews in the data. The many theme names retained from Study 2 possibly reflect the fact that it was by far the largest of the studies in terms of transcript length, relevant data by word count, and number of 'naturalness codes'. By the stage of the framework development at the end of that study, the themes and patterns in the data were becoming relatively clear and predictable. Perhaps as a result the wording was less ambiguous, and the nuances better expressed. By contrast the themes from Study 1 were more tentative and the framework from Study 3 was probably incomplete as discussed in Chapter 5.

The triangulated framework that resulted is shown in Figure 6.3 below. The themes relating to familiarity and full control have been shown in the centre of the framework to reflect the fact they underlie *any* safe and satisfying interaction in the car, but do not necessarily create natural-feeling interaction alone. The theme about comfort has been deleted because its role in naturalness was unclear, and sometimes contradictory. The physical, usability and future-focused humanlike intelligence themes are grouped together in three clusters reflecting the three higher order themes. These findings may now be carried forward to the final, quantitative, rating scale development study.

6.8 NATURALNESS FRAMEWORK WHICH EMERGED FROM TRIANGULATION

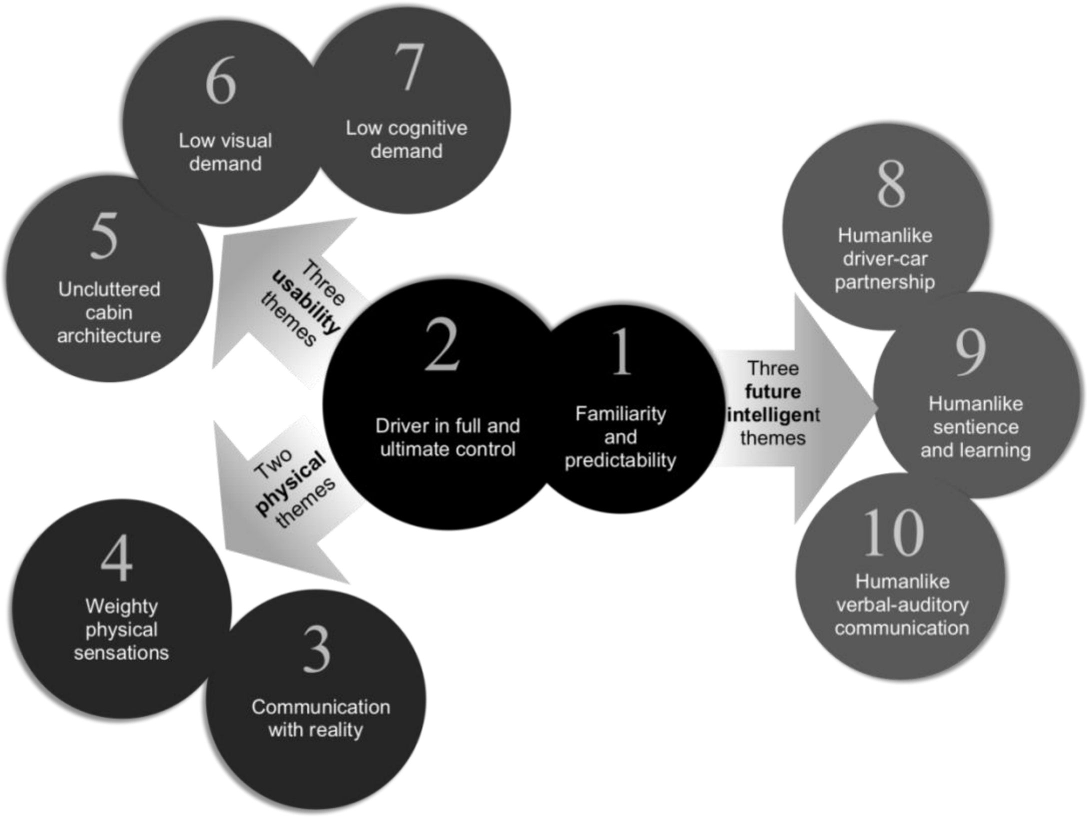


FIGURE 6.3 THE TRIANGULATED FRAMEWORK OF 10 CONSTRUCTS DERIVED FROM THE THREE QUALITATIVE STUDIES

6.9 CONCLUSION

Method Triangulation has provided 'trustworthiness' (Golafshani, 2003) to the three qualitative studies' results by using the principle of reflexivity and creating a transparent and bias-checked unified framework of natural-feeling secondary system interaction. The wording and layout of the naturalness framework has been finalised ready for conversion into questionnaire items. The triangulated naturalness framework now has 10 themes arranged as 'physical themes', 'usability themes' and 'future intelligent themes', with contributory themes about predictability and control.

CHAPTER 7

DEVELOPMENT AND TESTING OF A QUESTIONNAIRE-BASED INSTRUMENT TO EVALUATE PERCEPTION OF NATURALNESS IN SECONDARY CONTROLS

7.1 AIM

Qualitative research uses a naturalistic approach that seeks to understand phenomena in specific contexts (Golafshani, 2003). The previous three studies had all been qualitative because of the lack of existing understanding of driver-car naturalness. It was not yet known how much each of the themes might contribute to the construct as a whole, how the themes might correlate with each other, or how they would correlate with the word ‘natural’. A final, quantitative study was therefore considered necessary to interpret the qualitative results, giving statistical significance and strength to the various findings (Coolican, 2009).

Quantitative research measures causal relationships between variables (Denzin and Lincoln, 1998). Measurement scales are the foundation of scientific investigation (Stevens, 1946; Tal, 2015) and can provide an objective basis for critical decisions in industrial design (Tonetto and Desmet, 2016). Relevant and sensitive measurement tools will be needed in the future to optimise the design of automotive interfaces (Harris et al 2005). For the diverse results of this research to have maximum industrial application they needed to be translated into a means of measurement. This final study therefore aimed to create a valid, reliable, and rapidly deployable driver-car naturalness measurement scale for secondary car controls, for use with ordinary drivers.

7.2 RESEARCH METHODS

7.2.1 STANDARDISED QUESTIONNAIRES

Questionnaires are the most common instrument used in design research for obtaining quantitative data (Tonetto and Desmet, 2016). The process of taking a set of ordinary questions and converting it into a psychometrically valid and reliable "standardised" questionnaire involves a large number of participants answering a large set of often quite similar questions (‘items’) relating to the construct being investigated (Sauro, 2010b). These responses are then analysed to see which items correlate ‘strongly’ (considered in this research to be Pearson $r > 0.5$ with $p < 0.05$; Field, 2009) and elicit similar patterns of responses from subjects. Items with consistently ‘weak’ correlations (considered here to be $0.1 < r < 0.3$ or $p > 0.05$; Field, 2009) or unpredictable response patterns are eliminated to reduce the number of items down to a manageable set. Redundant or duplicated items (with similar wording and which correlate so highly that they must be effectively asking the same thing) may at this point be removed to make the final questionnaire shorter (Sauro, 2010b). The performance and correlation of all the remaining questionnaire items can then be compared to one another, to see if respondents are reliably evaluating one overall construct or several different underlying constructs (Weiner, 2007). Usually an exploratory factor analysis or principal components analysis (Field, 2009) is conducted to find out what components or factors underlie the measure in question. Corresponding ‘subscales’ may then be identified for the final tool. Finally, validation tests must then be conducted to check that the construct being measured is that which is intended (Weiner, 2007). If there is no related scale already validated in the literature this may be done by soliciting expert opinion in the form of criterion validation (predictive or concurrent; McDowell, 2006).

7.2.2 MEASUREMENT SCALES

Measurement scales are commonly based on the type of ‘standardised questionnaire’ described above. Measurement scales may seek to measure data at ratio, interval, ordinal or nominal level (Stevens, 1946) with ‘ratio’ being the most mathematically powerful. However, a complex and nuanced subjective quality like naturalness can at best be measured at ordinal level – meaning that only the rank order of responses is meaningful and can be known, but not their precise magnitude relative to zero nor the differences between those intervals (Field, 2009).

Few satisfactory measurement scales have been proposed for measurement of subjective qualities of automotive interfaces such as ‘usability’, ‘satisfaction’ or ‘pleasantness’; and none exist for naturalness of interaction. Harris et al (2005) developed a measurement scale to evaluate motor vehicle dynamic qualities (ride and handling) based on a scale used to measure aircraft handling qualities. Subjective rating scales are commonly deployed in gauging human work task demands more generally. The NASA-TLX subjective task load index is sometimes also used at the end of driving simulator studies (See Figure 7.1). The System Usability Scale (Brooke, 1996) is a well-known subjective rating scale for product usability widely used in Design Research.

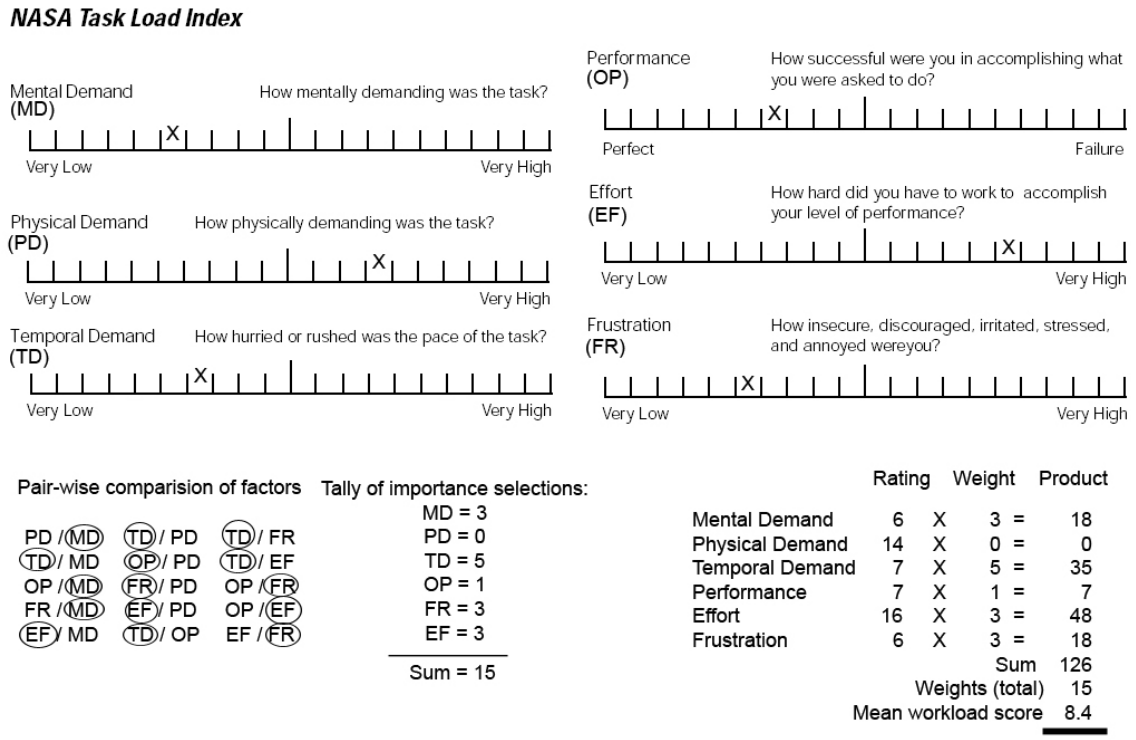


FIGURE 7.1 THE NASA TASK LOAD INDEX RATING SCALE FROM HART AND STAVELAND (1988)

Word based measurement scales are by far the most common and tend to be compact but risk biases in semantic interpretation of the question or anchor words themselves, or the measurement points (such as 'very' or 'somewhat'). Pictorial scales seek to reduce such linguistic biases where a 'feeling' is being measured across different cultures and groups, or where reading or comprehension may be a problem. The latter are most commonly used in healthcare assessment for measuring subjective feelings such as physical exertion (e.g. Marinov et al 2008) or nausea (Baxter et al 2011). However pictorial tools tend to take up more physical space (usually hardcopy, making the instrument more difficult to deliver) and may introduce visual biases. Highly nuanced perceptions can also be challenging to depict.

7.3 Study Objectives

The study was based broadly on the Nasa TLX development methodology described by Hart and Staveland (1988) using the seven-step measurement scale development process in Hinkin *et al.* (1997). These appeared to be efficient and well-practised ways of achieving the study aims. In summary the objectives were to:

1. Review the ten triangulated naturalness themes (and the verbal data that led to them) to extract all the possible underlying dimensions involved and produce a large number of items addressing those dimensions (the 'longlist');

2. Express these questionnaire items using 'domain specific' language of the users (Tonetto and Desmet, 2016) by pilot testing pictorial anchors and using guidelines from Hinkin (1998) to keep worded items short, specific and understandable;

3. Collect data by interviewing 80 to 100 drivers about a secondary control in their cars, while seated inside their cars (this being a suitable sample size for factor analysis);

4. Find the underlying factors behind the items in order to develop subscales;

5. Condense the large number of questionnaire items into a more manageable tool by keeping items which correlate highly with the concept as a whole;

6. Produce a final scale of a suitable length;

7. Validate the tool by conducting a suitable face validity test, and check that the quality being measured is indeed 'naturalness' by soliciting expert opinion on the 'naturalness' of various test controls (from a suitable test car or cars) and comparing it to the perceptions of ordinary drivers of the same controls.

7.4 Key Methodological Decisions

Before the study could start a number of key methodological decisions had to be made. It was observed that much of the verbal data on what constitutes natural-feeling driver-car interaction had been expressed in the form of qualitative polar opposite 'dimensions', rather than one-dimensional 'scales'. Participants were frequently observed to express *naturalness* by reference to what *unnaturalness* feel like – for example "an <u>un</u>cluttered central console" felt natural whereas a 'cluttered central console' felt unnatural (rather than neutral). Correct distinction between bipolar and unipolar is critical in affective measurement (Ekkikakis, 2013). Bipolar scales were therefore chosen for the study.

The benefits of a five-point bipolar scale have been frequently cited (e.g. Sauro, 2010). An odd number is generally preferred because many users wish to choose a mid-point or 'neutral' answer. Three-point bipolar scales do not allow the respondent the freedom to choose a 'somewhat' type response. Seven points may by contrast offer unwanted choices (Finstad, 2010) causing fatigue in questionnaires with more than about ten items (Sauro, 2010). Linguistic modifiers (such as 'high' and 'moderate') have been considered more suitable for eliciting subjective responses than numbered scales because people naturally describe experiences using such terms, rather than numbers (Hart and Staveland, 1988).

A pilot study attempted to portray the naturalness dimensions as pictorially anchored measurement scales with the help of a graphic artist and product designer. However, a word-based scale was ultimately chosen because it required rather less paper, and because pictures failed to represent the nuanced dimensions of naturalness adequately without introducing new biases relating to visual stereotyping, in the view of the lead researcher. Possession of a national driving licence is normally contingent on being able to read.

A survey of ordinal subjective measurement scales in common use in ergonomics (such as the SUS scale; Brooke,1996) identified 20 items as the maximum length of the final scale, with each individual user survey taking no longer than ten minutes to administer. Interviews with automotive interface professionals had also suggested a 15-minute maximum questionnaire length, using familiar and accessible language understandable by ordinary drivers, and a second more rapid version which might be used as a heuristic tool by professional designers themselves without the need for customer involvement.

Previous studies had been sparing in the use of the word 'natural' in order not to prejudice drivers' responses. However, if drivers might be asked *explicitly* how natural the control in question felt to use, there would be a separate measure of naturalness against which all the other *implicit* dimensions of naturalness could be compared. This appeared a highly useful opportunity. It would also help assess if the other items are in fact measuring 'naturalness' and not some different quality. This measure is referred to in the rest of this chapter as the '*explicit* naturalness' item (and in fact was the only item which used the word 'natural'). Any item which correlated poorly with it would be a cause for concern and a candidate for deletion. To avoid prejudicing respondents, the explicit naturalness item was asked around the middle of the questionnaire and was worded in a similar way to all the other items so as not to draw attention to it.

The final key methodological decision was at what 'level' (or granularity) of 'secondary control' the questionnaire items should be asked. There were three logical levels – the lowest level being an 'individual switch or dial' itself, the middle level being an entire ‘secondary system’ (which may have more than one switch, dial or display associated with it), or the highest level being ‘all secondary systems’ in any given car. A matrix was produced with likely advantages and disadvantages of each, based on learning from the previous three studies. These issues and implications were discussed with one other Brunel design researcher, for reasons of inter-rater reliability. The decision was to choose the middle (‘entire secondary system’) level – i.e. items asked about one secondary system *as a whole.* The items would be worded to apply to *every button switch screen or dial involved in that system.* Participants would be asked to *use* this system just before and during the questionnaire, in accordance with Contextual Inquiry practice (Beyer and Holtzblatt, 1997). As before the study would take place in participants’ own cars for ecological validity.

7.5 Sampling and Recruitment

Stratified Purposeful Sampling was chosen, to capture characteristics across a wide variety of participants (Patton, 1990). At the present time, different categories of car apparently display quite distinct interaction styles and secondary control systems, and a stratified spread was desired. The study also sought a variety of driver-automobile control scenarios include both routine scenarios (like adjusting the volume of a car radio) and some rarer more complicated scenarios (like calibrating a tyre pressure monitoring system). To assist with this, all respondents’ cars were classified according to the SMMT (2016) nine-class framework, with the aim of recruiting at least five participants per car type.

One tenth of subjects were recruited through university channels – the rest were recruited by appeals to the public through local neighbourhood websites such as Streetlife, community noticeboards and paid adverts in community websites, and through a live BBC Wiltshire radio feature. This was to reduce bias towards university students and males. Previous comparable automotive studies (e.g. Harris et al 2005) have shown a recruitment and self-selection bias towards males.

There is also no evidence that driver age has an effect on interaction preferences, but to exclude drivers lacking driving experience or with possible perceptual limitations, those aged under 25 or over 75 were excluded. To take part, respondents had to own or have regular use of a car, hold a driving license (of any nationality), and be fluent in English (because of the nuanced perceptions sought). Eighty-one drivers were recruited.

7.6 METHOD

7.6.1 DERIVATION OF THE LONGLIST OF ITEMS

All the possible underlying components of naturalness were extracted and compiled into a tentative longlist of specific, measureable, dimensional pairs of opposite adjectives or opposing statements. The wording of the originating participants was preserved where possible. This resulted in 63 bipolar dimensions possibly linked to perceptions of 'unnatural/natural-feeling' interaction. Each dimension was then expressed as an item with end anchors 'A' and 'B' (again mainly dictionary opposite adjectives but sometimes short sentences with opposite meanings) and five rating points 'Very A', 'Somewhat A',' neither A nor B', 'Somewhat B', and 'Very B'. This wording was based on a review of other subjective rating scales in use in the UK. To aid readability in this thesis, items are always presented with anchor 'A' predicted to correlate with 'unnatural-feeling' and anchor B with 'natural-feeling'. In practice half of the items were reverse scored during delivery to minimise order effects. *The word 'dimension' is replaced by the word 'item' from this point, because each bipolar 'dimension' formed a bipolar 'item'.*

7.6.2 REDUCTION OF THE LONGLIST OF ITEMS

A pilot test of four drivers rating one control each was conducted to test user acceptance of the 63 items and eliminate items that were perceived as too similar, confusing, or ambiguous. It was found to be generally well understood by drivers; industry-specific language caused most ambiguities (see Discussion, Section 7.13). The opportunity was taken to check face validity by asking these four drivers to comment on the items themselves and their perceived relevance to 'naturalness', after they had taken the test in their car. This reduced the number of items to 55. A second process involved discussion with an experienced automotive academic supervisor from Brunel University to further identify and reduce ambiguity, while ensuring each of the ten constructs was represented by a similar number of items. A target of four to five items for each construct was thought to be appropriate for this 'longlist' stage to avoid respondent fatigue. Finally, wording was simplified by discussion with three independent design researchers. Forty-four items remained.

7.6.3 QUESTIONNAIRE LAYOUT

The 44 items were typed into a legible three-page questionnaire for the researcher to read from and mark up the answers accurately. Use was made of shading and gridlines to reduce recording errors, but the answer boxes did not contain colour or numbers which might have prejudiced answers. The items were grouped into five clusters with similar themes. This was to assist participants in maintaining their concentration on similar themes rather than changing theme every item which might have caused mental fatigue. The 'explicit naturalness' item (i.e. the dimension 'Unnatural feeling'-'Natural feeling') was inserted twice; once near the beginning (Q7) and once at the end (Q41). The aim of repeating the item was to gauge the internal consistency and reliability of the data, in other words how stable 'explicit' naturalness perceptions were over the course of the interview – a possible criticism of a subjective construct like 'naturalness'. This made 46 items in total. The actual 'longlist' questionnaire format is shown in Appendix C.

7.7 PROCEDURE

The researcher travelled to a safe car park of the participant's choice, met them at their car, then sat in the passenger seat, with the participant requested to remain in the driver's seat. The format of the study was explained, and consent information provided before signature was requested. Brief demographic details (age, gender, car type and age) were collected for possible use in data analysis to explain patterns. A range of secondary systems were sought, from manual to intelligent or automated systems where they were present. The driver was asked to safely operate the system in question to familiarise him/herself with how it felt, before and during the interview. The 46 items were then asked, omitting the final five if there was no voice control. The system tested was recorded on the questionnaire.

FIGURE 7.2 PARTICIPANT ANSWERING ITEMS IN A TESLA SALOON CAR

7.8 Data Analysis

7.8.1 Manual 'within-case' data analysis

SPSS was used for the bulk of the analysis; but this software is not designed to analyse within-case correlations. Within-case correlations were considered to be highly useful for developing and validating the questionnaire, specifically comparing a subject's 44 individual item responses to their response to the item about 'how natural-feeling' that control felt. Therefore, each individual's rating for Q41 was *manually* compared to their response to every other item using their hardcopy questionnaires, a time-consuming task. No suitable pre-existing method could be found in the literature. Therefore, two logical scoring methods were developed. On a master sheet of all 46 items, for *each* item in turn, occurrences where a participant's rating matched *exactly* his or her rating for Q41 were recorded in one column; while occurrences of that item's rating being *within one scale point above or below* their Q41 rating were recorded in another column. Each person's questionnaire was analysed in this way giving a simple estimate of each item's correlation with the 'explicit naturalness' item. The items were then ranked in order of correlation in two ways. Firstly, in rank order of total instances of *exact correlation only* (Method 1). Secondly, by awarding one 'mark' for every direct match and a half 'mark' for every partial match within one scale point (Method 2). These results are given in Section 7.9.3.

7.8.2 Statistical analysis

SPSS was used for the main data analysis (treating the five-point scale data as ordinal – a reasonable assumption; Field, 2009). Firstly, Pearson's chi-square and phi tests were conducted to look for the effects of car type, intelligence of car system, participant age, and participant gender on naturalness ratings. Chi-square tests are suitable to look for relationships of this type between nominal and ordinal data (Field, 2009). The Pearson correlations between each item and every other item were calculated. These results are given in Section 7.9.4. In all the reporting below, significance is taken to be $p \leq 0.05$. In interpreting correlation coefficients, the following conventions (Coolican, 2009) are used: **.1 < r < .3 'weak'; .3 < r < .5 'moderate'; .5 < r < .8 'strong'; .8 < r < 1.0 'very strong'.**

7.8.3 Principal Components Analysis

Reduction in the number of questionnaire items is an important outcome of data analysis, parsimony may be achieved by explaining the variance of many interrelated variables in terms of a smaller number of underlying *factors* (Field, 2009). One of the most preferred ways of doing this is Principal Components Analysis (PCA; Field, 2009). 'Unrotated' factor solutions maximise the variance accounted for by the dominant factor by assuming all factors are unrelated and orthogonal. ‘Rotation’ tends to produce more ‘comprehensible’ and balanced factors with each item tending to load strongly on just one factor, rather than loading moderately on many (Field, 2009). A PCA was conducted with and without rotation. An initial analysis was run to obtain eigenvalues for each component in the data. Because of multicollinearity, six of the strongest correlating items (all Pearson $r > 0.7$) had to be eliminated followed by the 16 weakest items (those with $p > 0.1$ and ‘weak’ correlations). In fact the weakest correlating items tended to also have poor significance of $p > 0.1$. Once the matrix determinant was ‘acceptable’ ($> 1 \times 10^{-5}$; Field, 2009) there remained just 17 items in the PCA matrix, listed in Table 7.2. It is important to note that these 17 items did not represent any ‘shortlist’ for the final questionnaire, they were simply a selection demanded by the PCA process. Each valid factor iteration was checked semantically until there was common logic explaining each factor, giving coherent meaning *within* factors and distinctiveness *between* them (Patton, 1990). The underlying factors were helpfully named (Field, 2009) and tested for reliability using Cronbach’s alpha (with trial deletions) and the Kaiser-Meyer-Olkin (KMO) test for sampling adequacy (Field, 2009).

7.9 RESULTS

Eighty-one participants were tested (26F, 55M) each using **one** secondary system (i.e total of **81 secondary systems were tested**). Mean participant age was 47.4 years (SD=12.1; median age 43; the oldest was 68 and the youngest 25). All were fluent English speakers. Despite the appeals aimed equally at men and women, more men responded. The tests lasted an average of 21 minutes with the data collection taking 12 to 15 minutes. Of the car types used in the tests, 35% may be classified as premium/luxury (e.g. BMW 5 or 7 series), 17% small/medium premium car (e.g. BMW 1 or 3 series), 16% family car (e.g. Ford Focus/Mondeo), 16% large SUV (e.g. Range Rover), 14% small car (e.g. Ford Fiesta) and 3% sports car. Their mean age was 6.7 years old. The car types are listed in Table 7.1.

TABLE 7.1 BREAKDOWN OF CAR TYPES USED FOR THE TESTS

Car Type	Frequency	Percent
Family car	13	16
Luxury/premium car	28	35
Small/medium premium car	14	17
Small car	11	14
Sports car	2	3
large SUV	13	16
Total	**81**	**100.0**

The 81 systems tested comprised thirty-five different *types* of secondary system. The ten most common *types* of system tested were ventilation, heating and cooling systems (not automated climate control; 11%), windscreen wipers (11%), audio systems (10%), climate control systems (7%), electric windows (6%), trip computers/information systems (6%), indicators (6%), automatic/electric mirrors (4%), automatic headlights (4%) and keyless start systems (4%). The full list may be found in Appendix C. The car systems were also categorised for the analysis as either 'manual' (i.e. without automation but including electric assistance, e.g. electric seat adjustment), 'intelligent' (i.e. assistive and sensing systems which do not take control e.g. audible park distance sensing) or 'automated' (i.e. where the system senses *and* takes action e.g. climate control which maintains air temperature and quality automatically). These classifications were based on Bhise (2011) with reference to the NHTSA levels of automation (e.g. Flemisch et al, 2010). By this classification, 51% of systems tested were considered to be 'manual', 24% were 'intelligent' and 25% were 'automated'. This breakdown is shown in Figure 7.3 for interest but, as will be seen, had no significant effect on naturalness perceptions.

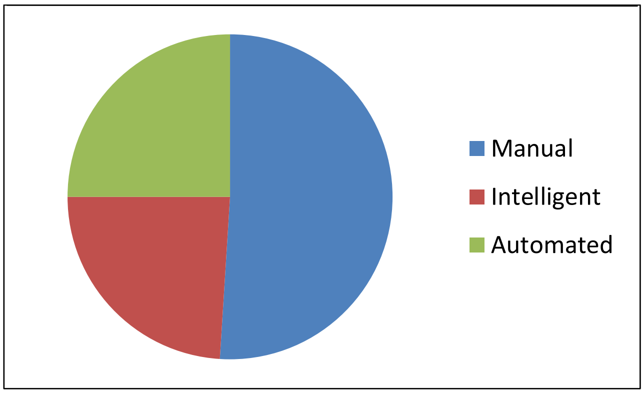
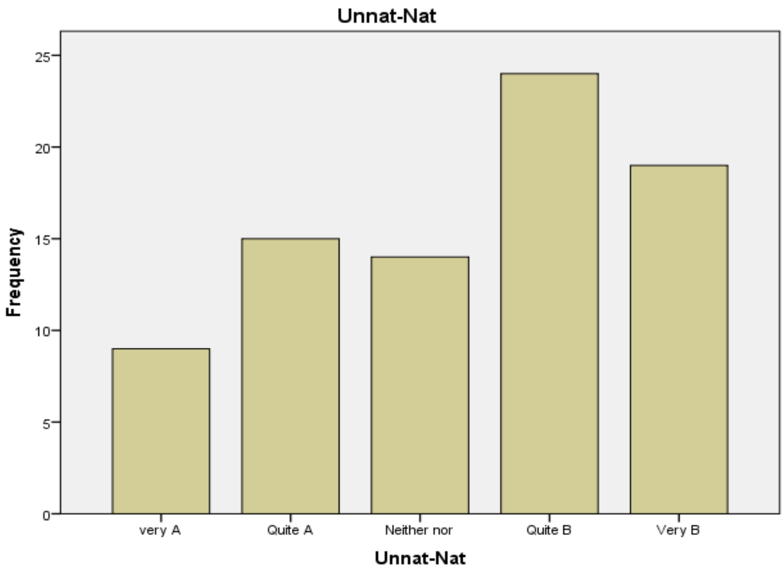


FIGURE 7.3 CLASSIFICATION OF CAR SYSTEMS ASSESSED IN THE TESTS

The voice activation items (Q42-45) could not be included in analysis because only one respondent answered them – this suggests a low uptake of current voice systems given that around one third of the cars tested had some voice input capability.

Although not intended to be a generalizable study, the 81 systems tested had an overall 'unnatural-feeling' to 'natural-feeling' profile as shown in Figure 7.4, with the left side of the chart ('Anchor A') being associated with 'unnatural-feeling' and the right side ('Anchor B') being associated with 'natural-feeling', and 'neither natural or unnatural' being in the centre of the chart. It can be seen that more systems were rated 'somewhat' or 'very' natural-feeling than any other category, and the least frequent rating was 'very unnatural-feeling'. In summary, perceptions tended to be skewed towards 'natural-feeling' (skewness -0.36; standard error of skewness 0.27).



N=81
Std. Deviation 1.326
Variance 1.758
Skewness -.362
Std. Error of Skewness .267

FIGURE 7.4 SUMMARY NATURALNESS PROFILE OF ALL THE SYSTEMS TESTED (UNNATURAL-FEELING TO THE LEFT, NATURAL-FEELING TO THE RIGHT)

7.9.1 STABILITY OF 'EXPLICIT' NATURALNESS RATING FROM Q7 TO Q41

Comparing the answer to the first explicit naturalness item (Q7) with the second (identical) item (Q41) showed a high degree of stability. In summary 70% of respondents answered the second naturalness item exactly the same as they did the first, while 30% had altered their answer by one point (up or down). No respondent altered their answer by more than one point. The same stability calculations on SPSS gave $r = 0.878$ and $\gamma = 0.941$ (both $p = 0.000$). This can be considered 'very strong' (Coolican, 2009). This data therefore suggests that naturalness was a stable perception over the course of the test.

7.9.2 EFFECTS OF CAR AND CONTROL TYPE ON NATURALNESS PERCEPTIONS

Effect of car type, driver age and gender on naturalness perceptions was investigated but no significant effects were found. Neither were any significant effects found between system type and naturalness ratings.

7.9.3 RESULTS OF WITHIN-CASE CORRELATION CALCULATIONS

Interpreting the within-case correlations at face value makes the assumption that the single item Q41 (the 'explicit' naturalness item) encapsulates the whole construct, which is unlikely. The results must therefore be viewed in parallel with the conventional statistics in Section 7.9.4. The ten strongest correlating items with 'unnatural-natural feeling' using within-case calculation Method 1 (see Section 7.8.1), listing strongest correlation at the top, and clarifying the meaning of the item in brackets the first time it appears only, were:

1. Artificial ↔ Real (i.e. communication between driver and car) Q9
2. Unhelpful ↔ Helpful (i.e. perception of car as helpful being) Q35
3. Unpredictably ↔ Predictably (i.e. car's response to commands) Q2
4. Counterintuitive ↔ Intuitive (i.e. interaction overall) Q27
5. For its own good ↔ for your own good (i.e. system is acting for whose benefit) Q38
6. Completely unclear ↔ Completely obvious (i.e. input action) Q30
7. Highly demanding ↔ Not at all demanding (i.e. mental demands) Q26
8. Difficult ↔ Easy (i.e. operation of the system) Q6
9. Rude ↔ Polite (i.e. perception of car as polite being) Q36
10. Highly incompetent ↔ Highly competent (i.e. perception of system's competency) Q37

The ten strongest correlating items with 'unnatural-natural feeling' using within-case calculation Method 2 (See Section 7.8.1), listing the strongest correlating at the top, were:

1. Counterintuitive ↔ Intuitive Q27
2. Artificial ↔ Real Q9
3. Unhelpful ↔ Helpful Q35
4. Unpredictably ↔ Predictably Q2
5. Difficult ↔ Easy Q6
6. Highly demanding ↔ Not at all demanding Q26
7. Rude ↔ Polite Q36
8. Completely unclear ↔ Completely obvious Q30
9. Highly incompetent ↔ Highly competent Q37
10. The car is fully in control ↔ You are fully in control Q3

Only the five weakest correlating items are reported in the two lists below because correlations increased markedly from 'weak' to 'moderate' after the fifth case in each list. In tandem with the correlations of each item with the measured construct (Section 7.9.4 below), it can be argued that the items below might be candidates for deletion from the scale, because they do not appear to be measuring the same quality as that which drivers interpret as 'natural/unnatural-feeling'. The five weakest correlating items with 'unnatural-natural feeling' using within-case Method 1 (weakest correlating at the top) were:

1. Digital ↔ Analogue (i.e. the look of the display) Q20
2. Forgets all your preferences ↔ Remembers all your preferences (system memory) Q40
3. Stays in the same position ↔ Returns to its original position (i.e. after use) Q17
4. Rigid ↔ Adaptable (i.e. (the ability of the system to adapt itself to circumstances) Q39
5. Passive ↔ Active (i.e. if the system is proactive or waits for input) Q33

Some showed weak negative (inverse) correlations. The five weakest correlating items with 'unnatural-natural feeling' using within-case Method 2 (weakest correlating at the top) were:

1. Digital ↔ Analogue Q20
2. Forgets all your preferences ↔ Remembers all your preferences Q40
3. Rigid ↔ Adaptable Q39
4. Lightweight ↔ Weighty (physical feel) Q12
5. Loose ↔ Tight (physical feel) Q13

7.9.4 Statistical correlations of the 39 items and the measured construct as a whole

The full correlation coefficient matrix is too large to be included (over 2,000 figures) but available on request. The following seven items had strong or very strong Pearson correlations with 'natural-feeling' Q41 (in fact all r >.7) and 'strong' correlations with almost every other item. All correlated as anticipated (i.e. positively not negatively), with acceptable significances. Listing the strongest correlating at the top:

1. Counterintuitive ↔ Intuitive Q27
2. Highly demanding ↔ Not at all demanding Q26
3. The car is fully in control ↔ You are fully in control Q3
4. Difficult ↔ Easy Q6
5. Unpredictably ↔ Predictably Q2
6. Artificial ↔ Real Q9
7. Only with careful thought ↔ Without careful thought (i.e. long term operation) Q31

The following eight items were eliminated from the final questionnaire because all had 'weak' to 'moderate' correlations combined with non-significant p-values. It should be noted that these are almost identical to the 'weak' items identified using the within-case methods in Section 7.9.3 above. The weakest is listed at the top:

1. Many locations ↔ one location (i.e. locus of control) Q5
2. Loose ↔ Tight Q13
3. Hollow ↔ Solid (i.e. physical feel) Q14
4. Slippery ↔ Tactile (i.e. physical feel) Q15
5. Stays in the same position ↔ Returns to its original position Q17
6. Digital ↔ Analogue Q20
7. Passive ↔ Active Q33
8. Forgets all your preferences ↔ Remembers all your preferences Q40

All the other 25 items fell somewhere between these two extremes – i.e. with acceptable *p*-values but only 'moderate' correlations. It would therefore be logical to prioritise the highest correlating items from the three lists in 7.9.3 and 7.9.4 for the final rating scale, providing they also contributed to one of the factors from the PCA (See Section 7.9.5 below). The 25 moderately correlating items might logically be used only if a factor was not served by the any of the strong correlating items. The eight weakest correlating items above arguably could not be retained at all, because they did not appear to be measuring the same construct as the other items, nor the perception 'natural/unnatural-feeling'.

7.9.5 Results of the Principal Components Analysis (PCA)

The optimal solution gave three underlying factors – this being the same whether Kaiser's criterion (eigenvalue > 1) or the scree plot method were used. These three factors in combination explained 63% of the variance. Table 7.2 shows the factor loadings after rotation which gave a more even distribution of variance between factors. The Kaiser-Meyer-Olkin (KMO) measure verified the sampling adequacy for the analysis (KMO = 0.882 ('very good' Field, 2009). All KMO values for individual items were above 0.8, well above the 'acceptable' limit of 0.5 (ibid). Bartlett's test of sphericity ($\chi 2$ (136) =811, $p < 0.001$) indicated that correlations between items were sufficiently large. Suppressing loadings below 0.4 (Field, 2009) and describing the items clustering on the factors gives the following solution:

Factor 1 represents *helpfulness and control*; items loading highly on this factor concerned 'helpfulness', 'politeness', 'competence', 'feedback', 'control', 'obviousness', and 'trust'. Of these, the strongest loading item was about 'helpfulness' ($r = 0.86$). After rotation, this factor accounted for 27% of variance. In all the valid iterations of the PCA, it did not prove possible to subdivide the primary factor into smaller logically discrete components. This indicates that when operating a car's secondary controls, drivers tend to perceive the apparently diverse concepts of 'helpfulness', 'politeness' and 'competence', as one and the same.

Factor 2 represents *strong communication and connection*. The item about 'instantaneous feedback' loaded highest ($r = 0.84$), followed by items about 'directness', 'connection', and 'eyes-free operation'. This factor appears semantically more focussed on a single issue than Factor 1. After rotation, this factor accounted for 21% of variance.

Factor 3 represents *logical location and form of controls*. After rotation, this factor accounted for 15% of variance. Items about 'logical location' and 'form reflecting function' loaded highest, followed by 'familiarity' and 'unclutteredness'. Again, this factor appears rather more semantically focused than Factor 1.

An alternative factor analysis method, principal axis factoring (Field, 2009) did not reveal a satisfactory logic behind its factors and had less even variance than the solution above.

7.9.6 Reliability analysis of the three factors and amendments

Cronbach's Alpha values (α) were all well over the minimum acceptable level suggested by Cronbach (1951) and Robinson et al. (1991) of 0.70.

Factor 1 had very high reliability ($\alpha = 0.917$) not requiring any item deletions because all α 'values if item deleted' were in excess of 0.895. The item loading lowest on this factor (the item about cluttered/unclutteredness) was deleted because it was the only one to have a loading of less than 0.5 and semantically had little in common with the other items. This left eight items.

Factor 2 had high reliability (initially $\alpha = 0.841$) with its six items all loading over 0.5. It was improved very slightly by deleting one item (about eyes-free operation) to $\alpha = 0.850$. Although not statistically necessary, this deletion did produce a more logically coherent and distinctive factor.

Factor 3 had acceptable reliability ($\alpha = 0.774$) with its five highest loading items. Deletions were attempted to improve reliability (to $\alpha > 0.8$) however the highest reliability was achieved by retaining all five items.

The factor analysis is summarised in the conventional manner in Table 7.2 overleaf.

TABLE 7.2 SUMMARY OF THE EXPLORATORY FACTOR ANALYSIS

Rotated Factor Loadings

Questionnaire Item	**Component**		
	1 Helpfulness and control	2 Strong communication and connection	3 Logical location and form
Unhelpful-Helpful Q35	.856		
Rude-Polite Q36	.783		
Incompetent-Competent Q37	.772		
Unclear-Obvious (input action) Q30	.654		.517
Car in control-You in control Q3	.648		
Not at all-Fully (trust) Q4	.563		
Delayed-Instant (response) Q16		.844	
Uncommunicative-Informative Q34		.765	
Indirect-Direct (communication) Q8		.706	
Weak-Strong (feedback) Q11		.623	
Divert eyes-Not need to divert eyes Q25		(.585)	
Artificial-Real (communication) Q9	.541	.579	
Trivial-Serious (feeling) Q10		.563	
Illogical-Logical (control location) Q19			.768
Shape reflects function-Does not reflect Q23			.640
Cluttered-Uncluttered Q21	(.486)		.586
Novel-Familiar (design and layout) Q1		(.419)	.535
Eigenvalues	**4.66**	**3.49**	**2.54**
% of Variance	**27.4%**	**20.5%**	**15.0%**
Cronbach alpha	**.917**	**.841**	**.774**

Extraction Method: Principal Component Analysis.
Loadings less than 0.4 not shown. Brackets indicate items later removed from a factor.
Rotation Method: Varimax with Kaiser Normalisation.
Rotation converged in nine iterations.

7.10 DEVELOPMENT OF THE FINAL MEASUREMENT SCALE

7.10.1 RELEVANT OBSERVATIONS DURING QUESTIONNAIRE USE

The questionnaire itself was to be a key output of the research. Although the study captured only numerical scores, drivers often reasoned their answers verbally. This had made it obvious where a question was ambiguous or misunderstood. Such observations made during the testing and refinement of the questionnaire were therefore noted. Perhaps as a result, all affected items were already deleted due to unacceptable correlations or significance. They may assist other researchers in modifying the questionnaire, however.

Confusion over terminology with specific HCI/ergonomic meaning

The intended meaning of words such as ‘digital’ and ‘analogue’ caused some confusion. Similarly, the terms ‘feedback’, ‘passive’ and ‘active’ often had different interpretations between subjects. ‘Rigid’ versus ‘adaptable’ was another example.

Confusion over common but inconsistently defined terminology

The word ‘direct’ sometimes caused confusion as to what interpretation of the word was intended (whether ‘quick’, ‘linked’ or ‘close coupled’). ‘Cluttered’ and ‘Uncluttered’ were terms that participants inconsistently defined between subjects.

‘Misperceptions’ about positions and memory of car controls

Much erratic answering followed questions about whether a control ‘stayed in the same position’ or ‘returned to its original position’, following use. A similar perception error seems to have also caused confusion over whether a control ‘forgets’ or ‘remembers’ preferences.

7.10.2 REDUCING THE NUMBER OF QUESTIONNAIRE ITEMS

The data and observations outlined above had suggested some items suitable for deletion. Research with industry professionals had indicated that the final measurement tool should be relatively brief to administer, no more than 20 items, and that it might come in a ‘normal’ and ‘short heuristic’ version. The lack of drivers using voice systems meant there were effectively 39 items at the start of the exercise. Using well-known design process techniques drawn from Hanington and Martin (2012) two such versions were developed.

Firstly, an independent psychology researcher from Oxford Brookes University assisted in an abbreviated form of the Affinity Mapping procedure (Hanington and Martin, 2012) to help reduce the items in the questionnaire from 39 to around 20 based on the statistics. First, criteria were agreed by negotiation with research supervisors and subject matter experts:

1. Normal version: 20 items maximum, some redundancy (at least one redundant item per subscale for greater reliability and robustness), duration 5-10 minutes, well-balanced with respect to items-per-factor, no weak correlating items, no moderately correlating items unless they load highly on an otherwise unrepresented subscale element. ‘Explicit naturalness’ item included discreetly.

2. Heuristic version: Five to ten items; capable of rapid delivery, no redundancy, only one item per factor (subscale), no weak or moderately correlating items unless a subscale element is not otherwise represented. ‘Explicit naturalness’ item included discreetly.

All the top correlating items (from within-case and Pearson correlation methods) were highlighted on a 39-item master sheet using coloured pens according to the origin of the correlation coefficient (whether within-case or Pearson). All weakly correlating items were deleted. The PCA factor numbers (1, 2, or 3) were marked on next to any item which loaded onto it, using asterisks for ‘strong’ loadings. This allowed patterns to be more visible – a key benefit of Affinity Mapping (Beyer and Holtzblatt, 1997). The master sheet is physically too large to be included but examples of selection criteria and key patterns identified include:

1. Items which correlated highly by **all three** calculation methods above:

- Unpredictably ↔ Predictably Q2
- Difficult ↔ Easy Q6
- Artificial ↔ Real Q9
- Highly demanding ↔ Not at all demanding Q26
- Counterintuitive ↔ Intuitive Q27

2. Items which correlated highly by **one** of the within-case methods **and** by Pearson *r* were:

- The car is fully in control ↔ You are fully in control Q3
- Completely unclear ↔ Completely obvious Q30
- Unhelpful ↔ Helpful Q35
- Rude ↔ Polite Q36
- Highly incompetent ↔ Highly competent Q37

The master sheet made redundant items simple to identify. Some redundancy was desirable – it can add to robustness and reliability by measuring important concepts in differently worded ways (Coolican, 2009) helping to even out differences between subjects’ linguistic perceptions. Excessive redundancy can however cause fatigue in respondents. Therefore, triply redundant items were deleted, retaining only the highest correlating items and two items per PCA factor. This left 15 items for the normal version which was considered an acceptable number. These are listed in Section 7.11.1.

The exercise also made it relatively clear which items to include in the heuristic version – logically this had to be any item which correlated highly (by at least one of the three methods of calculating) **and** loaded onto at least one of the three factors strongly. In effect, this simply eliminated redundancy. This reduced the instrument to ten items, considered an appropriate figure by the subject matter experts consulted. These are listed in Section 7.11.2.

It is recommended that the 'explicit' naturalness item is included in both versions as a reference point for possible future analysis, and as a reliability check. Based on experience of this study it is suggested that it is placed near the middle, and asked only once, because answers to this item did not appear to change significantly over the course of the test.

It is suggested that two items per subscale are inverted to minimise order effects, and that scoring is based on the same five-point scale as before (end anchors correlating with 'natural' and 'unnatural' and 'very' and 'somewhat' as intervals) because, as will be seen, face validity was demonstrated.

Should the data be analysed manually (scored at face value without statistical analysis) it is suggested that ratings be recorded as: '-2, -1-, 0, 1 or 2' and simply added up; this provides an immediate indication for whether a control in question is 'unnatural-feeling' or 'natural-feeling' because the midpoint 'neither unnatural nor natural' is equivalent to zero, a positive rating would mean 'natural-feeling' and a negative rating would mean 'unnatural-feeling'.

If computer statistical analysis is to be performed then a '0, 1, 2, 3 or 4' point rating system should be used (as in the analysis above) because most statistics software cannot compute negative numbers. This would mean the theoretical 'most unnatural-feeling car possible' would score zero and the scale could then be simply converted to a percentage. Because the midpoint would be two points out of four, then a 50-percent rating would become a useful neutral 'mid-point' reference in the scale. Any system scoring less than 50% being regarded on balance as more 'unnatural-feeling' than 'natural-feeling' and any system scoring more than 50% being more 'natural-feeling' than 'unnatural-feeling'. In due course, norms will need to be established to help contextualise ratings.

7.11 MEASUREMENT SCALE WHICH EMERGED FROM THE STUDY

7.11.1 NORMAL VERSION OF THE NATURALNESS MEASUREMENT SCALE

The three subscales and 15 items selected for the normal measurement scale were:

Subscale 1 – Helpfulness and control (maximum 32 points)

- Imagining the system as a person, it seems: Unhelpful ↔ Helpful
- Imagining the system as a person, it seems: Rude ↔ Polite
- The system seems: Highly incompetent ↔ Highly competent
- The car responds: Unpredictably ↔ Predictably
- When you use the control it feels like: The car is fully in control ↔ You are fully in control
- Operating the control feels: Difficult ↔ Easy
- Mentally the interaction is: Highly demanding ↔ Not at all demanding
- The interaction overall feels: Counter intuitive ↔ Intuitive

Subscale 2 – Strong communication and connection (maximum 12 or 16)

- The communication between you and the car feels: Artificial ↔ Real
- The control's response feels: Delayed ↔ Instant
- The car comes across as: Uncommunicative ↔ Informative
- Overall the interaction felt: Unnatural ↔ Natural

Subscale 3 – Logical location and form (maximum 12 points)

- The control is located: Illogically ↔ Logically
- The shape and action of the control: Does not reflect its function at all ↔ Closely reflects its function
- The input action required seems: Completely unclear ↔ Completely obvious

The maximum score possible is 60 points. Until norms can be established the mid-point of the scale (i.e. 30 points in this version) is assumed to be the neutral 'watershed' reference point. In some respects this is a logical assumption because the 'explicit' naturalness item has been referenced throughout analysis and questionnaire development. However, this one item cannot adequately capture the whole target construct. The actual questionnaire is shown in Figure 7.5 and uses graphic design and shading to enhance readability both for the administrator and the subject. Because of its importance to the findings, it is reproduced to almost actual size overleaf.

Name code:	Age:	Car used:	0	1	2	3	4	System used:	N= /60
		Attribute ‘A’	Very A	Somewhat A	Neither A nor B	Somewhat B	Very B	Attribute ‘B’	x5/3= % cf Q10 %
E	*[Example question only] The control’s action felt:*	*hard*						*easy*	*Score*
1	Imagining the car is a person, the system seems:	Unhelpful						Helpful	
2	Imagining the car is a person, the system seems:	Rude						Polite	
3	The system seems:	Highly incompetent						Highly competent	
4	The car responds:	Unpredictably						Predictably	
5	When you do the action it feels like:	The car is fully in control						You are fully in control	
6	Operating the control feels:	Difficult						Easy	
7	Mentally the interaction is:	Highly demanding						Not at all demanding	
8	The interaction overall feels:	Counter intuitive						Intuitive	
								HAC=	*/32*
9	The communication between you and the car feels:	Artificial						Real	
10	The control’s response feels:	Delayed						Instant	
11	The car comes across as:	Uncommunicative						Informative	
12	Overall the interaction felt:	Unnatural						Natural *SCC=*	*/12*
								N=	*/4*
13	The control is located:	Illogically						Logically	
14	The shape and movement of the control:	Does not reflect its function at all						Closely reflects its function	
15	The input action required seems:	Completely unclear						Completely obvious	
								LFF=	*/12*

FIGURE 7.5 THE NORMAL VERSION OF THE NATURALNESS MEASUREMENT SCALE

7.11.2 SHORT VERSION OF THE MEASUREMENT SCALE (10 ITEMS) FOR RAPID HEURISTIC DEPLOYMENT BY PROFESSIONALS

The ten items selected for the short version of the naturalness measurement scale were:

- Imagining the system as a person, it seems: Unhelpful ↔ Helpful
- The system seems: Highly incompetent ↔ Highly competent
- The car responds: Unpredictably ↔ Predictably
- Operating the control feels: Difficult ↔ Easy
- Mentally the interaction is: Highly demanding ↔ Not at all demanding
- The interaction overall feels: Counter intuitive ↔ Intuitive
- The communication between you and the car feels: Artificial ↔ Real
- Overall the interaction felt: Unnatural ↔ Natural
- The control is located: Illogically ↔ Logically
- The input action required seems: Completely unclear ↔ Completely obvious

The maximum score is 40 points. Any score more than 20 is on balance presumed to be ‘natural-feeling’ until norms can be established. The short version of the scale is illustrated in Figure 7.6 below.

Name code:	Age:	Car used for testing:	0	1	2	3	4	System used:	
Please answer the following questions honestly but without over-analysing. You are asked to rate the secondary system on various 5-point scales below in terms of how much you think it agrees with the ‘attribute’ in Columns A or B. The two attributes are roughly ‘opposites’. The mid-point represents ‘neither one nor the other’, the other points ‘very’ or ‘somewhat’. Keep repeating the action to remind yourself how it feels.		Attribute ‘A’	Very A	Somewhat A	Neither A nor B	Somewhat B	Very B	Attribute ‘B’	
E	*[Example question only] The control’s action felt:*	*hard*						*easy*	
1	Imagining the car is a person, the system seems:	Unhelpful						Helpful	
2	The system seems:	Highly incompetent						Highly competent	
3	The car responds:	Unpredictably						Predictably	
4	Operating the control feels:	Difficult						Easy	
5	Mentally the interaction is:	Highly demanding						Not at all demanding	
6	The interaction overall feels:	Counter intuitive						Intuitive	
7	The communication between you and the car feels:	Artificial						Real	
8	Overall the interaction felt:	Unnatural						Natural	
9	The control is located:	Illogically						Logically	
10	The input action required seems:	Completely unclear						Completely obvious	

FIGURE 7.6 THE SHORT VERSION OF THE NATURALNESS MEASUREMENT SCALE

7.12 Discussion

This chapter's various studies aimed to convert the research's largely naturalistic qualitative findings into quantifiable items and factors and find out the statistical relationships between them. This was in order to propose a measurement scale for naturalness of driver interaction with secondary controls and attempt to validate it. These aims have been achieved. The 15-item naturalness measurement scale should be a usable, valid, reliable and sensitive instrument where none exists at present. It takes around four minutes for a driver to rate one control, thereby meeting the targets for duration and number of questions suggested by the SMEs at the beginning of the research. Using five-point bipolar scales appeared logical and results suggest this was a valid approach, however future researchers may wish to repeat the study using unipolar scales (e.g. 'not natural – very natural') using three or five points.

The main correlation study, which asked 81 drivers to rate one control in their own cars using a 46-item scale, suggested that driver-car naturalness with secondary controls appeared to be closely correlated with perceptions of predictability, ease, helpfulness, politeness, full control, feedback, affordance, and competence. Perceptions of naturalness appeared stable over the course of a test. Many of these qualities were fairly predictable – being essentially carried forward from the qualitative triangulation and then upheld by statistical analysis. Unexpectedly however, almost all of the 'physical feel' items had to be excluded from the final framework and scale because of their weak correlations. This was unexpected because such issues had appeared to be highly salient in the first two studies. The only 'physical feel' item retained was 'delayed-instant', which could equally be related to 'competence' or 'feedback' rather than physical feel. This finding indicates that drivers may not be especially influenced by (or even perceptive of) switchgear feel when actually operating a control inside a *familiar* car, compared to when speaking about it, or operating a control *removed* from a car, or in an *unfamiliar* car. Studies such as Wellings et al (2008) presupposed the importance of switchgear feel, but (like in Chapter 4) tended to test this assumption inside unfamiliar cars or by laboratory 'bench testing' controls. However, they found drivers to be sensitive to switchgear *size* and *shape* (as did the present research).

Also unexpectedly, few of the 'mental and visual usability' items correlated significantly with naturalness. The previous two studies had suggested that these issues would be central. As suggested by the triangulation however, the first two studies appeared to exaggerate the contribution of physical usability because of their experimental conditions (equally the present study may have underrepresented usability issues because there was no concurrent driving task). Also deleted from the scale were the more descriptive items which attempted to classify perceptions as for example 'digital-analogue', 'rigid-adaptable', and 'passive-active'. This was apparently because of ambiguous interpretations leading to erratic answering. The other reason for weak (or inverse-weak) correlations of items appears to have been drivers' sometimes illogical perceptions of interactions as noted previously, which did not always reflect the objective reality. This particularly affected items about feedback or position of controls. In contrast, the items that asked drivers to think of the car as a 'sentient being' generally correlated strongly with naturalness and with the other items. Although many respondents had expressed surprise when asked such questions, once the intention was confirmed, few had trouble rating how 'rude' or 'helpful' a car was. As suggested by Reeves and Nass (1996), humans appear to instinctively treat intelligent machines much like human beings without necessarily being aware of it.

The principal components analysis suggested naturalness with secondary controls may be thought of as a combination of three reliable underlying factors related to: (1) helpfulness and control (2) strong communication and connection and (3) logical form and location. The high levels of multicolinearity revealed by the PCA together with the number of items correlating strongly with the 'explicit' naturalness item show the 'longlist' tool was probably measuring the same superordinate construct in multiple different ways, as was the intention at that stage. It also suggests that this quality was quite likely to be naturalness rather than some unrelated construct, and that drivers were answering the questionnaire fairly reliably. Factors 2 and 3 appear more semantically specific than Factor 1 – centring on single concepts. However, it could be argued that in Factor 1, 'helpfulness' and 'control' are in fact related. Helpfulness could for example be contributory to control because it suggests a status relationship where the driver is superior to the car. The same might be said of 'politeness' in Factor 1. Similarly, considering the apparent 'outlier' item (about eyes-free operation) in Factor 2 (about communication and feedback) it could be argued that a control which communicates very well, would generally permit the driver to look elsewhere.

This factor solution, with its high reliability and sampling adequacy, made it relatively straightforward to decide which items to include in the final succinct measurement scale. By comparing factor loadings with all the item correlation coefficients, any redundancy then became obvious. Item reduction was achieved largely by reducing redundancy while ensuring each factor was adequately represented by at least one item. The resulting subscale arrangement with individual subscale ratings may in future help manufacturers decide where strategic improvements might be made to a secondary system (existing or proposed) so that it has greater chance of being perceived as 'natural-feeling'. This may be important given the increasing cost proportion of electronics within secondary systems, the multiplicity of input modes and novel functionality now being offered, and the usability problems increasingly reported with such systems (e.g. Consumer Reports, 2016).

7.13 LIMITATIONS

The major limitation in this study was the reliance on subjects using their own cars for the test. This was deliberate - considered necessary to foster genuine, ‘naturalistic’, ‘situated’ and ‘occasioned’ interactions that – from the literature review – were assumed to be necessary (or at least sensible) in order to attempt to measure perceptions of naturalness. The disadvantage is that drivers were almost always long-term users of the car in question and may have been familiar enough with its secondary controls to make any interactions extremely ‘well-practised’ and perhaps even too ‘automatic’ to explain in detail. At least the static conditions inside a parked car arguably gave a little more potential and time to explore interaction perceptions more thoroughly compared to in a moving car with all the various concurrent dangers of the road. It may however have biased naturalness perceptions towards the ‘familiar’ rather than ‘novel’ and the overfamiliarity may have ‘blinded’ drivers to unnatural aspects of their cars’ controls which they had got used to over time.

In retrospect, a small companion test might have been carried out just after the main test, asking the same drivers to repeat the test in an *unfamiliar* car (using the equivalent type of control in both cars) and ask them to comment on the difference in feeling. This might have elicited some ‘unfamiliar’ ‘novel’ or less entrenched perceptions. However, in mitigation, when participants *had* used the researcher’s own car as a test vehicle in one of the early pilot studies, participants appeared rather nervous of damaging the car or activating the wrong control, or (more seriously) offending the researcher by saying anything negative about his car. This nervousness appeared to affect the honesty and clarity of results. By contrast, in their own cars drivers were observed generally to be relaxed and frank. It also took time to learn the operation and function of new controls in an unfamiliar car.

The static nature of the in-car test (parked rather than moving) may have detracted from the ecological validity of the study (but provided more control over environmental conditions than a moving car). Given the tests were mainly in public car parks, interaction stereotypes may have been biased towards parking scenarios and any strong emotions or memories that car parks evoked in the minds of subjects. In this study, each test involved 46 questions taking 12 to 15 minutes (the final scale will obviously be much shorter). It was observed that participants perhaps hesitated and thought about earlier questions much more than they did with later questions. It might have been preferable to invert the order of the question sets within the questionnaire to avoid participants possibly ‘overthinking’ initial sets of questions and ‘rushing’ the final sets. As it was, the *polarity* of the scale was inverted between participants (i.e. whether the polar anchors were ‘natural’ or ‘unnatural’ correlates) but not the *order* of the questions as a whole.

7.14 CONCLUSION

A questionnaire survey of 81 naturalistic driver-car-control scenarios was conducted, comprising 46 questions derived from the triangulated framework of ten themes and their possible underlying dimensions. From the results a statistically robust rating scale has been derived by means of within-case correlation calculations, Pearson correlation analysis and principal components analysis. Three underlying factors and 14 strongly correlating dimensions have been identified and used to create a succinct 15-item rating scale which satisfied all the usability criteria previously identified by subject matter experts and professionals. A shorter 'heuristic' version has also been suggested for use by automotive professionals. A scale is only valid if it measures what it claims to measure (McLeod, 2013). Procedures were therefore sought with which to test the content and criterion validity of the scale, which are described in the following Chapter.

CHAPTER 8

VALIDATION OF THE NATURALNESS MEASUREMENT SCALE

8.1 SCALE VALIDITY

A scale is valid if it measures what it claims to measure (McLeod, 2013). The *construct validity* of a scale is usually judged by convergent and discriminant validity. *Convergent validity* is when measures of constructs that should be related in theory, are shown to be related in practice, in a validation dataset (Patton, 1990). *Discriminant validity* is the opposite – essentially checking constructs that should not be related in theory, are in fact not related. However, naturalness has no rating scale in published literature, and no similar constructs could be found with rating scales. Neither did a scale exist for 'unnaturalness' or similar. Therefore, alternative means of validation were sought.

Criterion validation (usually *predictive* or *concurrent*) sometimes uses subject experts as the 'predictive' measure (Weiner, 2007) when there is no other suitable scale. Reliability may also be checked at this stage using representatives of the population from whom the scale is intended (McDowell, 2006).

Face validity is a less sophisticated test, which simply asks subjects (and sometimes experts) to check that a scale *appears* to measure what it purports to (Messick, 1989). It is a type of *content validity* (McDowell, 2006). Face validity may be conducted more rigorously by asking subjects to rate how 'essential' they perceive a particular questionnaire item to be, to the performance of the construct as a whole (such as in Lawshe, 1975) and calculating a coefficient from it. Face validity testing was conducted first, followed by criterion validation.

8.2 FACE VALIDITY TESTING

8.2.1 INTRODUCTION

Limited qualitative face validity exercises had already been conducted between the two studies described in Chapters 3 and 4, by presenting framework findings from the previous study to 25% of participants in the study immediately following it, for qualitative open-comment feedback. They were simply asked to assess to what extent they thought each theme *appeared to contribute to driver-car naturalness* and discuss with the group. A paper template of the framework had allowed space for qualitative comments and the results were later incorporated into the revised theme wording at triangulation stage (Chapter 6). However, these tests had considered only the early framework themes and not the validity of the final scale items, and had only consulted representatives of future scale users. In face validity involving both representatives of future subjects *and* subject matter experts (SMEs) can enhance its robustness (McDowell, 2006). Subjects may not always comprehend the diverse complex concepts involved in a target construct, while SMEs may not always be aware of the unpredictable perceptions of subjects.

8.2.2 OBJECTIVE

The objective was to conduct a rigorous face validity test involving representatives of the intended population (ordinary drivers) and SMEs.

8.2.3 METHOD

The Lawshe (1975) method of face validity estimation was considered to produce a simple yet rigorous measure of face validity, by asking if the concept measured by each item appears to be 'essential,' 'useful but not essential,' or 'not necessary' to the performance of the construct as a whole – followed by a four-point question about how suitable the whole scale appears to be for its purpose. This method was therefore chosen.

8.2.4 SAMPLING AND RECRUITMENT

The eight experts comprised five researchers at Brunel University (an automotive ergonomist, a HCI expert, a car mobile integration researcher, a vehicle vibration perception researcher, an inclusive designer, and an automotive biometrics researcher) and three automotive designers/researchers at automotive OEMs (one manager, one quality manager, one researcher). All were fluent in English. Six members of the group were selected for their automotive interface design experience specifically because they were representative of the type of industrial professionals who might use the measurement scale. The other two members of the SME group were chosen for their broader human-centred design research experience, particularly in measuring subjective human perceptions. They were consulted by email because of the large geographical distances between them and their busy schedules.

Thirty ordinary drivers who had also participated in the main study in Chapter 7, who had indicated that they would be prepared to help with future research, were also contacted to ask if they would also rate the 15 items. Sixteen responded and were interviewed in person. This was because they were in the same geographical area. All were fluent in English and met the criteria for ‘ordinary driver’ participants used in the previous chapter’s study.

8.2.5 PROCEDURE

A template of the 15-item scale was created with answer boxes for the Lawshe responses. Qualitative comments were encouraged in blank boxes. The test was first administered by email to eight SMEs and in-person to 30 ordinary drivers.

8.2.6 FACE VALIDITY FINDINGS

The experts' CVR scores are shown in Table 8.1, with scores above zero indicating that *most* experts agreed that an item was *essential*.

TABLE 8.1 EXPERT FACE VALIDITY RATINGS OF THE 15 ITEMS (LAWSHE METHOD) N=8

Questionnaire Item	**CVR score**	**'Positive rating' % [i.e. 'essential' *or* 'useful']**
1. Imagining the system as a person, it seems: Unhelpful ↔ Helpful	-.75	100
2. Imagining the system as a person, it seems: Rude ↔ Polite	-.50	62.5
3. The system seems: Highly incompetent ↔ Highly competent	-.25	100
4. The car responds: Unpredictably ↔ Predictably	1.00	100
5. When you use…: The car is… ↔ You are fully in control	.75	100
6. Operating the control feels: Difficult ↔ Easy	.50	100
7. Mentally the interaction is: Highly…↔ Not at all demanding	.25	100
8. The interaction overall feels: Counter intuitive ↔ Intuitive	.25	100
9. Communication between you and the car feels: Artificial ↔ Real	.25	100
10. The control's response feels: Delayed ↔ Instant	-.50	100
11. The car comes across as: Uncommunicative ↔ Informative	.25	100
12. Overall the interaction felt: Unnatural ↔ Natural	.25	100
13. The control is located: Illogically ↔ Logically	-.75	100
14. Shape/action of the control: Does not…↔ Closely reflects its function	-.50	87.5
15. Input action required: Completely unclear ↔ Completely obvious	.25	100

Overall, seven of the eight experts rated the scale as 'very suitable for its purpose' and one expert as 'extremely suitable for its purpose' which was interpreted as a positive face validity finding. It can be seen that five items scored negatively (that is fewer than half of the experts thought they were 'essential') these being items 1, 2, 3, 10, 13, and 14. The percentage of 'positive' face validity ratings, that is the combination of 'essential' and 'useful' but excluding 'not necessary', is shown in the third column of Table 8.1. This shows that when the category of 'useful' was included in 'positive' appraisals, only items 2 and 14 were not 'positively' appraised by *all* the experts (these items were about 'politeness' and 'control shape'). However, these experts were not aware of the correlations of these two items with naturalness, and their statistical significance in the subscale factors, calculated from the 'learning' dataset. Earlier in this chapter it had been observed that both 'politeness' and 'control shape' items had high correlations with the overall concept, and with nearly every other item. The scale sought to measure drivers' perceptions of naturalness, not those of experts. Therefore, both were retained.

The ordinary drivers' face validity CVR scores are shown in Table 8.2. Comparing the two tables, it can be seen that these ordinary drivers rated the scale items more positively than the experts for face validity, for all but two of the 15 items.

TABLE 8.2 ORDINARY DRIVER FACE VALIDITY RATINGS OF THE 15 ITEMS (N=16)

Questionnaire Item	CVR score
1. Imagining the system as a person, it seems: Unhelpful ↔ Helpful	0
2. Imagining the system as a person, it seems: Rude ↔ Polite	0
3. The system seems: Highly incompetent ↔ Highly competent	.25
4. The car responds: Unpredictably ↔ Predictably	.88
5. When you use the control it feels like: The car is…. ↔ You are fully in control	.53
6. Operating the control feels: Difficult ↔ Easy	1
7. Mentally the interaction is: Highly demanding ↔ Not at all demanding	.75
8. The interaction overall feels: Counter intuitive ↔ Intuitive	.88
9. The communication between you and the car feels: Artificial ↔ Real	.63
10. The control's response feels: Delayed ↔ Instant	.25
11. The car comes across as: Uncommunicative ↔ Informative	.50
12. Overall the interaction felt: Unnatural ↔ Natural	1
13. The control is located: Illogically ↔ Logically	.38
14. The shape & action of the control: Does not…↔ Closely reflects its function	.75
15. The input action required seems: Completely unclear ↔ Completely obvious	.63

No items had negative CVR scores, which suggests all had positive face validity apart from items 1 and 2 which were rated as 'neutral' – i.e. the same number of drivers rated them as 'essential' as not. This was a scale intended for use on ordinary drivers. The two items, about 'helpfulness' and 'politeness', were retained because of their strong Pearson correlations and major contributions to the primary subscale shown by the data analysis in the previous chapter (factor loadings both above 0.75). Although half the participants may not have appreciated the 'face value' role of helpfulness and politeness in naturalness, the

81 naturalistic driver-car-control scenarios had shown strong associations. It is possible that the general positivity of the ratings may be a result of 'people pleasing' social biases (Dvorsky, 2013). These were rather unavoidable given that the participants at this late stage would possibly have been under the impression that they were 'rating' this researcher's core thesis. However, in mitigation, the wording of the test delivery had urged participants to be honest in their appraisals, and had stated that negative appraisals of questionnaire items would be equally as useful to the thesis as positive appraisals of questionnaire items.

8.3 Criterion validation test

8.3.1 Introduction

A literature review had identified *predictive/concurrent 'criterion' validation* (using subject matter experts as the 'criterion' measure of the construct of naturalness) as the only feasible means of external scale validation (Weiner, 2007) because there were no other suitable measures with which to validate concurrently. Using expert opinion to validate a new scale predictively is a common strategy in linguistic and health quality measurement. For example, Fischer (1984) asked 53 native speakers of French to independently rate 18 written texts by non-native learners for their 'communicative quality', in order to validate a new objective points-based measurement scale designed to measure the same quality. A rigorous predictive criterion validation study was therefore designed, as the second validation test.

8.3.2 Objective

The objective was to ask several experts to rate independently the 'naturalness' of various secondary systems (using the same overall scale anchors/points as the measurement scale, but *not* using the full instrument). The mean of their ratings would then be compared to the ratings of a sample of the ordinary car driver population using the 15-item instrument (answering questions about the **same** systems in the **same** car as the experts). 'Agreement' type statistics such as t-tests could then be used to test for difference in the findings.

8.3.3 Method

For logistical reasons, it was necessary to have all the test systems located in the same test car, in order to be able to compare ratings across subjects and experts efficiently. The test car chosen was a 2006 BMW 530d Touring SE. This vehicle was chosen because it was a high specification 'premium' brand car (featuring a colour HUD and adaptive headlights, for example), about the mean age of a UK car at the time of the test (which was 7.8 years old; SMMT, 2016). Since secondary systems generally currently exhibit a wide range of ‘intelligence’ and ‘naturalness’ (Norman, 2005) it was important to test a corresponding range of systems. The test car exhibited a range of secondary systems which may be perceived as ‘manual’, ‘intelligent’, and ‘automated’. The car’s secondary systems were also anticipated to exhibit the full range of ‘natural’ to ‘unnatural’ perceptions in operation, based on contemporary Consumer Reports articles (e.g. 2014) and the motoring press. Likely perceptions of all the test car’s secondary systems were ‘predicted’ by one automotive researcher and one ordinary driver, and categorised using a six-cell matrix of ‘manual *or* intelligent *or* automatic’ against ‘natural *or* unnatural’.

Six secondary systems were then selected, one each from the six categories ‘likely to be perceived as’: (1) ‘manual’ and ‘unnatural’, (2) ‘manual’ and ‘natural’, (3) ‘intelligent’ and ‘unnatural’, (4) ‘intelligent’ and ‘natural’, (5) ‘automated’ and ‘unnatural’, and (6) ‘automated’ and ‘natural’. These were:

1. A six-way electric **seat adjustment** system operated from a control mounted on the side of the seat base (featuring separately shaped ‘seat base’ and ‘backrest’ controls with four-way movements directly mapped to the seat parts. Hidden during normal use, this control is compromised by adjacent memory input/recall controls which can change the mode, stopping further adjustment or suddenly moving all parts of the seat to another driver’s memorised position. It has no display. The only feedback is the movement of the seat itself. The motors have a delay in operation around half a second. *This system was anticipated to be perceived as ‘manual’ and ‘unnatural’.*

2. Electric four-way **steering wheel adjustment** via a single steering column mounted ‘joystick’-type control with four degrees of freedom (up, down, forward, back) mapped logically to the wheel’s movement. It has no display and the only feedback is the movement of the steering wheel itself. There is no perceptible delay. *This system was anticipated to be perceived as ‘manual’ and ‘natural’.*

3. A multifunction controller operated **GPS** with digital map and traffic information. It has two graphical displays (the HUD and a large console screen) as well as voice inputs and outputs. The controller knob has basic haptic feedback (e.g. number of detents adapts to number of options on the current menu) and has a stiff feel with five degrees of freedom (forward, back, right, left, down) but few explanations of system possibilities. An early version of the BMW iDrive system, with up to two seconds’ delay, it *was anticipated to be perceived as ‘intelligent’ and ‘unnatural’.*

4. **Wiper** system with rain sensitivity setting. The wiper moves eccentrically to sweep a large screen area. Its feedback consists of a single LED to signify mode, and the visible

movement of the wipers themselves. The rain-sensitivity mode is cancelled at the end of each journey, always defaulting to manual mode. The lever is touch sensitive and sprung but always springs back to the same position. *It was anticipated to be perceived as 'intelligent' and 'natural'.*

5. A fully automated **climate control** system. User input is restricted to temperature only, everything else such as fan speed and air conditioning are controlled by computer. These settings can be partly adjusted by going into the iDrive computer system via four sub-menus. Automatic mode is always on by default, it may be disengaged but only fan speed and temperature can be manually controlled. Its display consists of three LEDs, and hard-to-see white dial markers. *This system was anticipated to be perceived as 'automated' and 'unnatural'.*

6. An adaptive automatic **headlights** system which turns lights on and off according to ambient light conditions, and controls high beam automatically according to oncoming traffic. It aims the headlamp beam around corners, taking a feed from the steering angle. In the event of oncoming lights, the high beam always defaults to 'dipped'. Feedback is by the road lighting itself and by two icons in the main binnacle. *This system was anticipated to be perceived as 'automated' and 'natural'.*

The six systems' controls are shown in Figure 8.1 below.

8.3.4 Sampling and recruitment

The five experts selected were two senior automotive interface designers and three automotive ergonomists, three based at a large UK car company and two from Brunel University. They provided the '*predictive*' measure of naturalness. Sixteen ordinary drivers were recruited from Brunel University research offices in the School of Engineering and Design and community websites in Wiltshire, in the same way as described for the main study. They were the '*representative participants'* for the future scale.

8.3.5 Procedure

To make the test fair, consistent and reproducible, it was essential to equip both the experts and the drivers with the correct knowledge of the functioning of each system but not to prejudice them with positive or negative opinions about it. A neutral overview was therefore given of each of the six systems' basic functionality and settings, based on the test car's instruction manual. Participants were then requested to operate and explore each system before answering any questions about it, as follows.

The test car was first driven to a safe car park near the place of work of each expert. They were seated in the driving seat and asked to rate the six systems (and any associated input controls and outputs) on a single five-point scale from ‘very unnatural’ to ‘very natural’, according to their ‘professional opinion’. To assist in their judgment, they were given only this research’s definition of naturalness (see Chapter 2). The 16 ordinary drivers were asked to rate the same six systems on a different day, but using the naturalness measurement scale (normal version) instead. Thus 96 driver ‘tests’ were conducted in total (each 15 items long). The experts did not meet the ordinary drivers. For both tests the car was kept clean and all personal possessions were removed to minimise prejudice and maintain a ‘neutral’ setting in which participants felt comfortable in honestly appraising the various aspects. The order of the systems was rotated by one each test, to address possible order effects.

FIGURE 8.1. THE SIX SECONDARY SYSTEMS TESTED IN VALIDATION: (1) SEAT ADJUSTMENT, (2) STEERING WHEEL ADJUSTMENT, (3) GPS, (4) WIPERS, (5) CLIMATE (6) HEADLIGHTS.

8.3.6 Criterion validity test findings

Tests took an average (mean) of 24 minutes including familiarisation time. This means that on average, each test took four minutes. This is just below the lower duration target suggested by the SMEs. The five experts' ratings were compared to the 16 ordinary drivers' ratings in various ways. Figure 8.2 shows the 16 drivers' *mean* rating for each system plotted against the five experts' *mean* rating for each system, as a scatter plot.

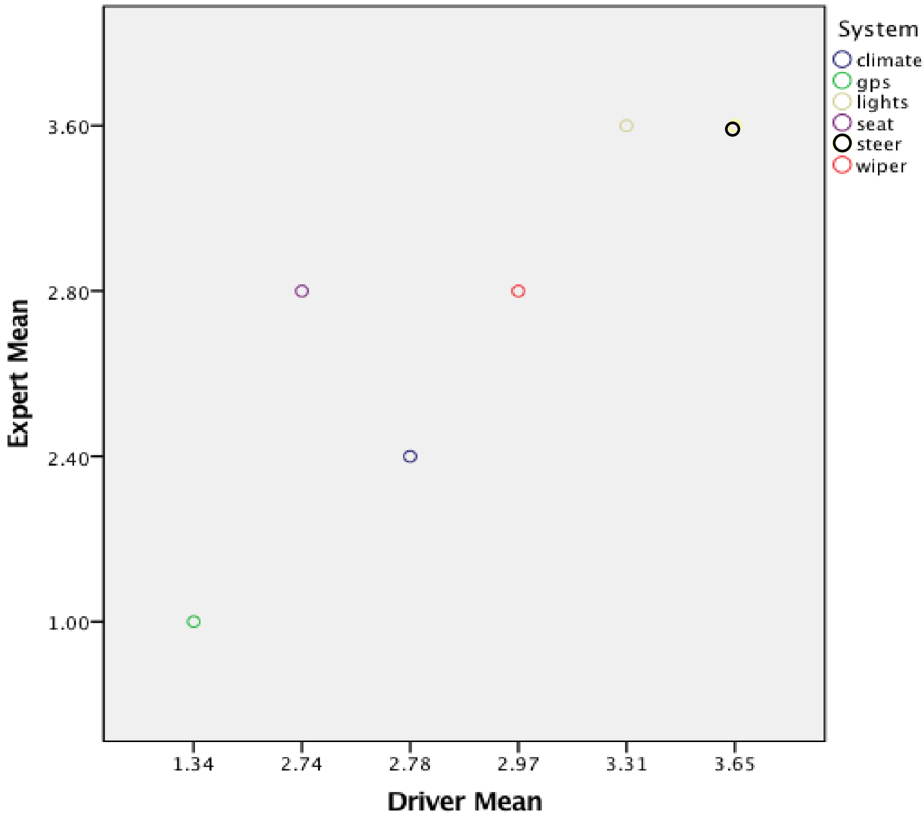


Figure 8.2 Scatter plot of Drivers' mean naturalness rating for each of the 6 systems tested, against Experts' mean rating for the same system

It can be seen that there is a reasonably close correlation between the two groups' mean ratings. Systems that were rated low on naturalness by drivers were also rated low by experts, etc. SPSS software was then used to look for more rigorous 'agreement' between the two groups' results treating the scale data as interval. Firstly, Spearman Rho correlations between the two groups' naturalness ratings were calculated, and summarised in Table 8.3.

TABLE 8.3 SPEARMAN RHO CORRELATIONS BETWEEN DRIVERS' MEAN RATINGS AND EXPERTS' MEAN RATINGS FOR EACH SYSTEM PAIRWISE

Spearman Rho test			Driver Mean	Expert Mean
	Driver Mean	Correlation Coefficient	1.000	.829*
		Sig. (2-tailed)	.	.042
		N (i.e. of test systems)	6	6
	Expert Mean	Correlation Coefficient	.829*	1.000
		Sig. (2-tailed)	.042	.
		N	6	6

The Spearman Rho test is a 'relative' measure of agreement which considers only the rank order of ratings (not their magnitude). It showed a correlation of 0.829 between the drivers' naturalness scale ratings and the experts' ratings (the assumed proxy for the 'valid' construct at this point). Spearman Rho correlations over 0.7 are considered to be 'good' (Field, 2009). This suggests that the measurement scale put the six systems in a very similar rank order of naturalness as did the experts. It cannot demonstrate that their ratings were the same. Since both groups were effectively using the same five-point scale, a Pearson correlation was also conducted. This was 'very strong' (r = 0.961, $p < 0.002$) although the sample size was too small to base meaningful conclusions from this. A more detailed overview was obtained by comparing the pattern of overall individual scores of the experts with those of the drivers. This was done using an Independent-Samples t-Test to test for difference between the two groups with the null hypothesis being that there was no significant difference. The results are shown in Tables 8.4 and 8.5.

TABLE 8.4 INDEPENDENT T-TEST DRIVER/EXPERT GROUP STATISTICS COMPARING MEANS

System	Group	N	Mean	Std. Deviation	Std. Error Mean
GPS	Driver	16	1.34	.54	.13
	Expert	5	0.80	1.30	.58
Climate Control	Driver	16	2.78	.74	.19
	Expert	5	2.20	.84	.37
Seat Adjustment	Driver	16	2.74	.66	.16
	Expert	5	2.80	.45	.20
Wipers	Driver	16	2.97	.55	.14
	Expert	5	2.40	1.34	.60
Headlights	Driver	16	3.31	.46	.12
	Expert	5	3.40	.55	.24
Steering Wheel Adjust	Driver	16	3.65	.43	.11
	Expert	5	3.60	.55	.24

The Levene test (Field, 2009) suggested equal variance could be assumed for the climate control, seat adjustment, headlights and steering wheel but not the GPS or wipers. This indicates that the expert and driver rating *variances* were not significantly different for four of the six systems.

TABLE 8.5 INDEPENDENT SAMPLES T-TEST FOR EQUALITY OF MEANS BETWEEN DRIVERS' NATURALNESS SCORES AND EXPERTS' NATURALNESS RATINGS

	Levene's Test		t-test for Equality of Means				
	F	Sig.	t-value	df	Sig. (2-tailed)	Mean Diff.	Std Error Diff.
GPS	5.72	.027	**1.37**	19	.383	.538	.392
Climate Control	.04	.851	**1.48**	19	.355	.578	.392
Seat Adjustment	.91	.353	**-0.20**	19	.846	-.063	.317
Wipers	13.64	.002	**1.42**	19	.639	.571	.402
Headlights	.76	.396	**-0.38**	19	.709	-.094	.248
Steering Wheel Adjustment	1.73	.205	**0.21**	19	.834	.050	.233

For a two-tailed test with *df*=19, t-values should be in excess of 2.093 to demonstrate significant *difference* in means to $p < 0.05$ (Field, 2009). In other words, a t-value of less than 2.093 would indicate that the null hypothesis holds, i.e. there is no statistically significant difference in means between the two groups. Hence the analysis suggested that **drivers using the naturalness scale did not produce significantly different naturalness ratings to the experts**, with respect to all six car systems tested. Using Tables 8.4 and 8.5, for example comparing the mean (*M*) Steering wheel adjustment ratings (Row 5) M_{expert} = 3.60, SD = 0.55; and M_{driver} = 3.65, SD =0.43; the difference between means was not significant, t(19) = 0.21, $p < 0.05$, two-tailed, the effect size (r = 0.05) was 'small' (Field, 2009). However, the effect size was not 'small' for all the systems, which may be related to small sample size (Coolican, 2009). The effect size was small (below 0.1) for the headlights and steering wheel adjustment, but moderate (0.3) for the GPS, climate control, seat adjustment and wipers. This perhaps slightly weakens the conclusion that the expert ratings and driver ratings were not significantly different for all systems.

Considering briefly the magnitude of the ratings (only tentatively because of the small sample size), the steering wheel adjustment showed the lowest mean difference between groups (almost identical in magnitude). The highest mean difference (just over half a scale point) was found in the case of the climate control. Overall, three systems (the GPS, the climate control and the wipers) were rated more natural by drivers than by experts, with means different by about half a scale point. The other three systems were rated very similarly in magnitude by both groups (the means were within one tenth of a scale point).

8.4 Reliability

Reliability of the scale was not tested directly. While not a direct measure of reliability, the relatively small standard deviations of the drivers' ratings (around half a scale point) suggests that drivers were rating the same systems reliably (between subjects). Another proxy for reliability (within subjects) was obtained by comparing *each* drivers' calculated naturalness rating for *each* system tested (i.e. that system's rating obtained from the scale) with that driver's corresponding answer to the 'explicit naturalness' item (i.e. for the same system). A Pearson correlation was then conducted on these 96 data pairs (16 drivers x 6 systems). A high correlation would arguably indicate some degree of reliability (and also some degree of face validity, because it would show that the overall construct measured by the scale was similar to the concept drivers understood by the words 'unnatural-natural feeling'). The Pearson correlation was found to be 'very strong' ($r = 0.909$, $p < .001$) indicating reliability of the scale. Further studies will be needed on test-retest reliability between subjects.

The reliability or consistency of the ratings *between* experts was markedly less consistent than that between the drivers, varying by as much as two scale points (out of five) between experts. These industrial automotive experts appeared to have strong feelings about certain aspects of usability in the systems tested, and were noticeably more bimodal in their ratings than the participants (i.e. judging the same systems either harshly or favourably). This suggests that SMEs are not reliable predictors of driver-car naturalness when a single expert alone rates a single system. The heuristic short form of the scale may be beneficial to help them; alternative study designs are suggested in the next section.

8.5 Discussion

In combination with the ecological validity of the construct and scale development, and the basic checks on reliability permitted by the 'explicit naturalness' scale-check item, the procedures described above were considered to be an acceptable initial validation procedure, but one which further studies will need to build upon. Face validity appeared to be acceptable. The results also suggested that the naturalness scale can provide a valid measure of naturalness when used by ordinary drivers, compared to the professional opinions of experts. There was evidence of reliability from the correlations with the 'check' naturalness question and the consistency of ratings between drivers for the same systems. This also suggested that the scale was rating a similar construct as that which is understood by the words 'natural/unnatural-feeling'. Clearly this only applies to the six systems chosen on the test car; different test cars and more systems will be needed to fully validate the scale.

The inconsistency of expert ratings is perhaps cause for concern but again suggests the need for a rating scale designed for *ordinary* drivers. Perhaps the lack of expert consensus was due to their subject matter expertise itself - a consequence of their specialisation and different professional interests. Using the short *heuristic* form of the scale may prove to be a more reliable and 'accurate' way for experts to assess naturalness, compared to simply asking them their opinion. However, this has not yet been tested. Two alternative validation study designs are suggested to overcome the problem of the unreliability of ratings between experts. Firstly, a *panel* of experts might *together* assess a system's naturalness by discussion and consensus, which might provide a more considered and reliable measure than averaging out the personal opinion of each expert in isolation as was done here. Secondly, a group of *ordinary drivers* (rather than experts) might provide the predictive measure of naturalness. This could be for example asking one group of drivers to assess various systems' naturalness using only the *definition* of naturalness given in Chapter 2, while a second group of drivers rates the same systems using the naturalness measurement scale. Agreement and difference would again be tested as above.

Using a car with which subjects were *unfamiliar* appeared challenging for experts and drivers alike, as compared to the main study which had used familiar systems in participants' own cars. Mappings and functionality were sometimes unfamiliar and needed to be learned quickly. In addition, some functionality was harder to demonstrate during particular tests because of environmental conditions. For example, the adaptive headlights were much easier to demonstrate in low light conditions, and the 'intelligent' wipers easier to demonstrate when it was raining (which in fact it was on both test days hence its inclusion in the results). Another weakness of the test was that the car was stationary and there were no hazards. This meant that the issues of 'eyes off road time' and mental demands were perhaps underrepresented in naturalness ratings, because there was no concurrent primary driving task for participants to be distracted *from*. The branding, detailing and engine sound of the test car could not be concealed (a BMW estate car with a three-litre six-cylinder diesel engine) therefore stereotyping and biases may have occurred as to the 'likely' performance of secondary systems in a car like that – Chapter 3 had suggested that drivers sometimes do apply car brand stereotypes (such as 'German efficiency') to cars' secondary systems.

Finally, although *discriminant validity* was not explicitly tested, the six systems selected did in fact display a wide variety of naturalness ratings from ‘very unnatural’ to ‘very natural’, as was hoped. Further validity testing will be needed on *pairs or groups* of car systems (with the same function but different levels of anticipated naturalness) to ascertain if the scale can discriminate significantly, and helpfully, between them. Further testing will also be required to establish *norms* (Coolican, 2009) for the rating scale, to better explain and contextualise future results from it.

8.6 Conclusion

The naturalness measurement scale was validated using face validity rating methods, and a rigorous criterion validity measure which compared 16 ordinary drivers’ ratings of six test systems against six experts’ naturalness ratings of the same systems in the same car. The tests suggested acceptable face/content validity (judged by both experts and ordinary drivers) with half the items rated ‘very high’ (CVR $\geq$ 0.75) by at least one group, and nine out of fourteen items rated ‘high’ (CVR $\geq$ 0.5) by a sample of intended recipients of the scale. There was acceptable predictive validity between drivers rating naturalness using the scale and experts rating naturalness using their professional opinions. Ordinary drivers rated systems in the same ‘order of naturalness’ as the experts, and with very similar magnitude and variance. Validation will benefit from future work in a wider range of test vehicles, but with the limited resources available the test suggested the scale is fit for purpose and may now proceed to more thorough reliability testing, discriminant validation, and alternative validation testing, using a more reliable predictive measure against which to test the scale.

CHAPTER 9

OVERALL CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

Following a wide-ranging literature review on a topic about which little was known academically, four different studies were designed to address an exploratory research question regarding what constitutes natural-feeling driver-car interaction with secondary controls and systems, and how it might be measured. Five qualitative human centred exploratory methods were adapted to the challenging confines of the automobile in order to study the nuanced interactional perceptions of ordinary drivers, who normally operate their secondary controls alone, silently and apparently not always fully consciously. Those methods were Contextual Inquiry (with ethnographic interview), exploratory design workshops (flexible modelling, focus groups, and 'Think Aloud' analysis), and ethnography (participant observation including a substantial future fiction study). Each study was analysed according to established academic procedures using multiple additional independent researchers to maximise objectivity. Method triangulation was then employed to identify, and attempt to minimise, the effects of experimental biases, in order to draw together all the findings into a unified framework. Finally, a substantial quantitative study was performed which first converted the framework findings into bipolar dimensions, and then into 46 five-point scale items, for a large questionnaire study. Eighty-one drivers answered this, each rating one secondary system in their own car while seated within that car and using its associated controls, to maintain high levels of 'naturalistic' validity.

From these naturalistic scenarios, correlations and underlying factors were estimated, and subsequently the 'world's first' driver-car naturalness measurement scale was developed. This succinct instrument consists of 15 items, all of which were found to strongly and significantly correlate with natural-feeling interaction, and contribute to one or more of its three highly reliable subscales. The measurement scale was then validated to some extent using two fairly rigorous methods addressing content and criterion validity.

The methods used, and their adaptations, prioritised ecological validity, largely recruiting and studying 'ordinary driver' participants outside universities. This was considered essential in research which set out to explore and measure the nuanced 'situated' and 'occasioned' phenomenon of naturalness. This gives the findings rather more validity than the driving simulator studies which dominate automotive interface literature (Reich and Stark, 2016).

WHAT MAKES NATURALNESS DISTINCT FROM CLASSICAL USABILITY AND HUMAN FACTORS DESIGN?

Study after study, the findings suggested that while the phenomenon that was observed and measured *incorporates many* elements of classical 'usability', the concept of driver-car naturalness is **greater than simply 'good usability' or good human factors practice**. For example, referring to the rating scale study and the component dimensions derived from it (reproduced below as Section 9.1), dimensions 3-8, and 10-12 can all be paired to recommendations in human factors or 'usability' literature (such as Norman, 2012). However, the two *highest loading* dimensions (numbers 1 and 2 according to the factor analysis in Section 7.9) do *not* have a direct usability counterpart - helpfulness, politeness (as well as instantaneousness) are unique to the *naturalness* of driver-car interaction and suggest an important contribution to the field. Neither are these two or three qualities particularly prominent in any of the more recent interaction design paradigms of say 'emotional design' or 'pleasurable design' which in comparison feel rather superficial with their references to plastic smiles and digital hugs. Even if the reader were to doubt the validity of the scale, qualities related to the two or three unique concepts in question appeared in themes in <u>all</u> the other studies – demonstrated by the number of references to 'humanlikeness' 'helpfulness' (and 'connection') throughout this thesis.

An example of a secondary control which might be rated as both usable *and* natural, is the rain sensing automatic windscreen wiper system on the author's 2007 BMW 5-series (illustrated in Figure 9.1). It will be described in detail because it illustrates qualities of helpfulness, humanlike algorithms, coherent metaphor and instantaneousness over and above its more predictable usability and functional qualities. Specifically, this may be demonstrated in the operation of its sensitivity selector, a click-detent thumbwheel on the lever itself which is moved up and down as required. It may be easily operated eyes-free because the position can be easily felt. It is logically and conveniently located (on the familiar column stalk which operates only the wipers). It gives feedback through vibration, sound, position and resulting action (four modes). It is, by most definitions, usable. But it goes further. The metaphor is coherent and clear (Celentano and Dubois, 2014) - one may speak of a person or animal becoming 'increasingly or decreasingly sensitive or aroused', and the general metaphorical convention in this context would be 'up' for increased sensitivity and 'down' for decreasing sensitivity. Furthermore, its operational behaviour is humanlike and helpful – demonstrating a natural understanding of the operator's *human needs* in the likely *situations* and *occasions* it will be used. Because when the sensitivity wheel is clicked up any amount, the wipers *always* perform a confirmatory single sweep of the screen, unrequested, regardless of how hard (or even if) it is raining. In doing so the system has predicted, usually correctly in the author's experience, that increasing the sensitivity of the rain sensor is not an intellectual or detached action likely to be performed in a dry weather or abstract situation. It tends to only be adjusted when the driver feels the *wipers are not coming on enough*. In doing the unrequested wipe, it gives feedback both that the system has both received *and* understood the driver's intent. By contrast, if the same wheel is clicked *down* a notch (to make it less sensitive) the wipers do not perform any such confirmatory wipe. This is perhaps counter to 'usability' conventions of confirmation and consistency. If the system does a sweep to confirm a sensitivity increase, why does it not do one to confirm a decrease? However, this feels *natural* because the system has predicted, again correctly in the author's experience, that the only situation that a driver would turn *down* the wiper sensitivity is if they felt the wipers were *coming on too much* (and perhaps streaking, juddering or otherwise damaging the rubber blade). Doing a similar confirmatory sweep in *this* scenario would therefore likely be perceived as annoying, incompetent, or 'unhelpful'. That theoretical scenario might therefore be rated 'usable' but 'unnatural'. This control was in fact tested in Chapter 8 as part of the validation and ratings tended to be in the 'somewhat natural to very natural' range – but subjects were using the control for the first time and for a very short period (not long enough for any rain to increase or decrease in severity).

FIGURE 9.1 RAIN-SENSING AUTOMATIC WINDSCREEN WIPER SYSTEM ON 2007 BMW 5-SERIES

Similarly, 'up' and 'down' buttons for continuously variable adjustment of volume, fan speed or cabin temperature may be rated as 'usable'. That such controls are seen in so many cars presumably means they are NHTSA safety compliant, have satisfied the many usability checklists OEMs tend to have in place for interface design, and presumably have not caused an excessive amount of customer complaints. Yet drivers in the present studies tended to rate such controls as less than fully natural. It may be theorised this was because they had to divert their eyes to choose the right button, or they could not tell by feel where the desired control was, or which one was ‘up’ and which was ‘down’. Perhaps there was nothing to ‘grab onto’ or hold – this they could not take the weight or tension out of their arm muscles while operating it (a variant of the ‘appropriation for comfort’ theme seen in Chapter 6). There is no natural, anthropological or skeuomorphic metaphor for ‘down’ meaning ‘tap something on the left’ nor up meaning ‘tap something on the right’. Indeed, this is not a universal convention nor followed by every car tested (sometimes ‘up’ was on the left). These are all ‘unnatural’ qualities as suggested in the studies that make up this thesis. By contrast, round *dials* for volume or temperature tended to be rated as very natural – perhaps because of historical/skeuomorphic associations with volume knobs from home audio systems, or the dials on home thermostats. The metaphor of ‘energy up’ being ‘turn clockwise’ and ‘energy down’ being ‘turn anticlockwise’ itself possibly derives from the centuries old movement of clock hands.

Touchscreens also meet many usability criteria but were often perceived as unnatural, as was the electronic handbrake described in Chapter 6 which may by some definitions also be described as ‘usable’. The typical electronic handbrake has an LED to indicate state, it requires very little force to operate, it is located near the drivers’ resting hand position – all qualities associated with usability. However, it is *unnatural* by this thesis’ criteria: its state is not clearly communicated at a glance, it does not have the reality-based connection and feeling of ‘tension’ that a conventional handbrake has when applied or released; electronic handbrakes may also be unpredictable in action and automation, it can feel like the car is in control, and the interaction may feel artificial and delayed (all unnatural qualities according to this thesis’s findings). Thus, unnatural perceptions appear to have led to an unnatural rating.

Naturalness in a car may have **other benefits** not tested in this thesis: it can be proposed that there may be **safety** benefits to controls which can be operated without looking or thinking, or as a result of reduced frustration or anger modulation (as in Harris and Nass, 2011) or from reduced mode confusion (Sarter and Woods, 1995). There may be safety benefits from greater perceptions of ‘connection’ to environmental hazards like rain, ice, standing water (Walker et al, 2006). Multimodality may have safety benefits of not leaving controls activated when not intended (such as indicators left on accidentally on a motorway). Arguably there may be commercial benefits of naturalness too, such as increased pleasantness and satisfaction leading to repeat purchase or more favourable consumer reviews (Fleischmann, 2007).

To recap, the research questions were:

RQ1. What are the component characteristics and dimensions of natural-feeling interaction between ordinary drivers and automobile secondary controls and in what circumstances does it tend to occur? How do these dimensions correlate with each other and with drivers’ interpretations of the word ‘natural’?

RQ2. What factors underlie the construct of natural-feeling interaction between ordinary drivers and automobile secondary controls?

RQ3. By measuring and rating drivers’ naturalness perceptions of various interactions with their own cars can a valid reliable measurement scale for driver-car naturalness be developed that is also relevant to the future?

RQ4. How might driver-car interaction still feel natural in future more intelligent or highly automated cars?

The achievements and conclusions of the research are summarised below according to which parts of the research question were answered, in the logical order RQ1 to RQ4.

When interpreting Pearson correlation coefficients, the following conventions (Coolican, 2009) are used: $0.1 < r < 0.3$ is ‘weak’; $0.3 < r < 0.5$ is ‘moderate’; $0.5 < r < 0.8$ is ‘strong’; and $0.8 < r < 1.0$ is ‘very strong’ (Coolican, 2009). When discussing content (face) validity, the following definitions are used: CVR ≤ 0 is ‘unacceptable’; CVR 0.01-0.24 is ‘low’; CVR 0.25-0.49 is ‘moderate’; CVR 0.5-0.74 is ‘high’; CVR 0.75-1.0 is ‘very high’ (Lawshe, 1975).

9.1 The 14 Component Dimensions of Naturalness (RQ1)

The most logically complete answer to the first part of the research question will be achieved by considering the 14 bipolar items which formed the final validated naturalness measurement scale, and tracing each ‘item-dimension’ (i.e. naturalness characteristic) back through the four studies showing how it arose in each. The reasons for inclusion in the final framework and scale can then be justified. *Although the scale has 15* items*, one of them is a ‘check’ item explicitly asking how un/natural the system feels; there are only 14* dimensions*.*

1. Imagining the car as a person, the system seems: Unhelpful—Helpful

The theme of natural-feeling interaction being perceived as ‘helpful’ arose in all studies. It arose in the interview study in themes of ‘proactivity’ and ‘assistance’. In the exploratory workshop study it manifested as ‘partnership’. The participant observation study had many observations which confirmed the importance of ‘helpfulness’ in naturalness, despite the obvious differences in dictionary definitions of ‘naturalness’ and ‘helpfulness’ (which are likely to have contributed to the lower face validity ratings for this item). In the statistical analysis of the questionnaire study however, the item *unhelpful-helpful* correlated consistently strongly with the construct as a whole, and with the ‘unnatural/natural-feeling’ item. It was also the strongest loading item on the primary principal component from the questionnaire study, Factor 1 ‘helpfulness and control’.

2. Imagining the car as a person, the system seems: Rude—Polite

The theme of ‘politeness’ was a theme that arose independently in all studies in slightly different forms. In the interview study it manifested itself within the themes of ‘assistance’ and ‘subservience’, and was implicit in the various ‘humanlike’ behaviour themes in the interview study and the exploratory workshop study. The future-fiction participant observation study showed the strong importance of ‘politeness’ in automotive technology acceptance and possible naturalness perceptions of intelligent cars. Again, while the dictionary definitions of ‘politeness’ and ‘naturalness’ have little in common, the statistical correlation was clear in the case of car secondary controls. The apparent semantic differences between ‘rude-polite’ and ‘unnatural-natural’ are likely to have contributed to the low face validity ratings by experts and drivers, the lowest of all the items.

3. The system seems: Highly Incompetent—Highly Competent

The theme of ‘competence’ was subtly referenced in several themes from the interview and exploratory workshop studies (for example ‘intelligence’ and ‘predictability’). The participant observation studies gave the item *highly incompetent-highly competent* particular emphasis however. This finding in particular may offer OEMs strategic guidance on spending the increasing proportion of build-cost associated with secondary system electronics – it is not ‘intelligence’ or ‘breadth of ability’ which correlated with naturalness but mere ‘competence’. A cheaper, well-engineered, simple, ‘competent’ system may be perceived as more natural than a highly intelligent system with multiple features, settings and modes.

4. The car responds: Unpredictably—Predictably

The theme of ‘predictability’ arose as ‘expectation’ or ‘familiarity’ in every study, and the item *unpredictable-predictable* exhibited consistently high correlations with naturalness and most of the other items. Although it may be a more general theme that underlies *any* kind of satisfying interaction, not just natural-feeling interaction, the statistical link could not be ignored. The item also had ‘very high’ and ‘high’ face validity, when judged by ordinary drivers and experts respectively. It is one of several items that may be linked by a superordinate theme of ‘clarity of design metaphor’ (see Point 15 below).

5. It feels like: The car is fully in control—You are fully in control

The theme of ‘control’ arose explicitly in every study. It was the strongest single theme in the interview study (according to content analysis), a central theme in the exploratory workshop study, and was strongly implied in some of the interpretations of participant observations. This was one item where finding the bipolar ‘opposite’ was not clear cut, and therefore may warrant further investigation. Feasible ‘opposites’ of the statement ‘I felt fully in control’ (which was the dimension’s positive anchor, as originally extracted from the theme) could have been ‘I felt fully *out* of control’ or the more unipolar ‘I *did not* feel in control’. Expressing the opposite as ‘the *car* felt fully in control’, as it is in the final scale, may appear excessive. However, it appeared to encompass the perceptions of many participants when they spoke about the cars they owned (which had a mean age of 6.7 years and therefore tended not to feature many automated systems in which the car might seem to be ‘in control’). It also correlated more strongly than other similarly worded items in the ‘longlist’ questionnaire. In future however, when more drivers possess cars with competent automated systems or with full self-driving capabilities, this choice of wording may negatively prejudice against them. Naturalness ratings for this item might consistently be low even if the automation is otherwise perceived as ‘natural-feeling’. In the current analysis, however, the item as written was the third highest correlating item overall with good expert-rated face validity. Together with the ‘helpfulness’ item it formed the basis of the strongest factor, Factor 1 ‘helpfulness and control’.

6. Operating the control feels: Difficult—Easy

Multiple themes from every study suggested that ‘easy-feeling’ interaction feels natural in the automobile. In initial studies this concept was present in the themes about ‘usability’ and ‘low cognitive demand’. The three themes that together may be described as concerning ‘easiness’ (this one and the two below) were very prolific according to content analysis and strong in the correlation statistics (without being so similar that they were redundant). Therefore, keeping them as three distinct items in the final measurement scale appeared justified (Items 6, 7 and 8). The *difficult-easy* item unsurprisingly had very high combined face validity (adding the experts’ CVR scores to the drivers’) – the highest of any item.

7. Mentally the interaction is: Highly demanding—Not at all demanding

The evidence for this theme was as above. The inclusion of this item was merited on the basis of high correlations by every method used. The semantic link between ‘undemanding’ and ‘natural-feeling’ interaction is relatively clear and predictable. When judged by ordinary drivers, this item had very high face validity.

8. The interaction overall feels: Counter intuitive— Intuitive

This item was included on the basis of its high correlations, which were the highest of any item. It was the highest correlating item of all, by two out of the three methods of calculation used. The semantic link between ‘intuitive-feeling’ and ‘natural-feeling’ is clear and predictable. When judged by ordinary drivers this item also had high face validity.

9. The communication between you and the car feels: Artificial—Real

There was a ‘reality-based’ theme found in every study. In the interview study this was the theme about ‘direct connection’ with the car and a second related theme about ‘physical feel’. In the exploratory workshop study the theme was called ‘communication with reality’ reflecting the naturalness preference for automotive interactions that felt ‘grounded in reality’ and ‘not like a computer game’. The theme wordings differed between the first two studies partly in response to feedback in one of the small inter-study face validity exercises, revealing some anachronistic ‘mechanical’ connotations with the phrase ‘direct connection’. The item *artificial-real,* largely formed from drivers’ own words, was the highest correlating item with the *words* ‘unnatural/natural-feeling’ according to manual correlation. It loaded moderately on *two* of the three factors in the principal components analysis (Factors 1 and 2) – one of the few items to do so after rotation. Again, the semantic links between the dimension ‘artificial-real’ and the construct ‘unnatural-natural’ are relatively clear.

10. The control’s response feels: Delayed—Instant

The theme of ‘instantly responding’ secondary controls being perceived as natural-feeling, was revealed by all the studies apart from the interview study (the interview study was in fact the only study that did not require drivers to operate real car controls during data collection so this might have lessened the salience of ‘instantaneousness’). Whereas most of the final scale items were highly subjective (yet correlated predictably and strongly) this was an example of an item that actually had an *objective* element (i.e. measurable time delay) yet appeared to be inaccurately or subjectively answered by drivers nevertheless. During the questionnaire data collection, time delays that were obvious to the researcher were not always perceived as such by drivers, who often rated systems with an obvious input-response time lag as ‘instant-feeling’. This is an example of the sometimes nonlinear, unpredictable responses of ordinary drivers described in previous discussion sections. The item *delayed-instant* correlated highly with naturalness despite that. The item *delayed-*

instant was the single strongest item in Factor 2 ‘strong communication and connection’ in the principal components solution to the questionnaire study findings.

11. The car comes across as: Uncommunicative—Informative

The theme of ‘informative machine communication’ (commonly known as feedback) was observed in the interview study’s definition of the ‘direct connection’ and ‘vehicular usability’ themes, and noted in all other studies’ data. The participant observation study particularly suggested that a natural-feeling dashboard should give ample and appropriate feedback. Specifically, it was suggested in the discussion section of that study that any single secondary system might exhibit *multimodal* feedback in at least two sensory channels (implying also multiple *redundancy*). This may be particularly ‘natural’ because road noise, visual demands and sensory overload mean that one human processing channel is sometimes temporarily compromised. The item *uncommunicative-informative* loaded highly on Factor 1 ‘helpfulness and control’. This item is one of only two items retained in the final scale, with Item 10 above, which possibly (but not necessarily) relate to ‘physical feel’, the others all having been deleted for reasons of ambiguity and poor correlation.

12. The control is located: Illogically—Logically

The theme of ‘logical location’ of controls was captured in the theme of ‘vehicular usability’ in the interview study, and within the theme of ‘low cognitive demand’ in the exploratory workshop study, when its data was reviewed. In the participant observation study, it was observed that drivers often semi-consciously located or checked frequently-used controls using their ‘feel’, shape or location/position alone. This suggested that *logical, proximal* and *familiar* locations may all contribute to naturalness. However, drivers also rated certain *far-away* and *unfamiliarly located* controls as ‘natural-feeling’ if they were rarely used, or hidden for safety reasons (for example, a bonnet release lever is typically ‘hard to reach’ so that the bonnet is not raised accidentally at speed, but drivers would still describe it as ‘natural-feeling’). That is why the final item avoided use of anchor wordings like ‘easy to find’ or ‘nearby’ in favour of ‘illogical/logical location’. All studies suggested that if the designer’s choice of control location (for example ‘hidden away for safety’ or ‘nearby for convenience’) is *clear and logical*, it may be perceived naturally (see ‘clarity of design metaphor’ discussion in Point 15 below). The item *illogically-logically located* was the strongest loading item on Factor 3 in the Principal Components Analysis, ‘logical location and form’. When judged by ordinary drivers this item had very high face validity, whereas experts rated it as only moderate.

13. The shape and movement of the control: Does not reflect its function at all—Closely reflects its function

The theme of 'control shape suggesting action' was noted in the participant observation study and in the exploratory workshops. This property is similar to the concept of 'affordance' (Norman, 2013). During driver observations, the naturalness of affordance had appeared to be further enhanced if the movement of the control itself mirrored the consequent motion of the secondary system controlled by it. For example, the windscreen wiper lever's arc rotation in some ways suggests the rotation of the wiper arms themselves. Although correlations were moderate overall in the questionnaire study, the item was retained because of its high loading on Factor 3 'logical location and form'. Again, this item may be linked to the overarching theme of 'clarity of design metaphor' (see Point 15 below).

14. The input action seems: Completely unclear—Completely obvious

The theme of 'obvious control action' in naturalness was noted in the participant observation and the exploratory workshop studies. It correlated highly in the manual within-case correlations and loaded on two of the three factors in the principal components solution – Factor 1 'helpfulness and control' and Factor 3 'logical location and form'. Again, this property is similar to Norman's (2013) concept of 'affordance'. When judged by ordinary drivers this item had very high face validity, whereas experts rated it only as moderate.

15. Summary and Higher-Level Patterns Observed in the 14 Items

In summary fourteen descriptive dimensions, each expressed as approximately polar-opposite scale items, were identified qualitatively and then suggested statistically to be contributors to the target construct of natural-feeling interaction with cars' secondary systems. Around one third of the items could reasonably have been **predicted** – those relating to dictionary synonyms of 'naturalness', such as 'intuitiveness' and 'predictability' for example. Unexpectedly however, perceptions of 'physical feel' appeared statistically to be largely irrelevant to perceptions of naturalness, although it could be argued that the two items relating to 'feedback' and 'instant response' do concern physical feel. Together with 'logic of location' and the 'affordance' items (about control shape and movement), **physical/architectural items** may be considered to constitute just one third of the scale. The items which attempted to measure less tangible relationship dynamics and **the car as a semi-social/intelligent being**, contributed more than anticipated – making up the remaining one third of the scale items.

Common to many of the 14 items, especially Items 1, 2 and 3, was an implied concept that natural-feeling interaction is **'humanlike interaction'**. In other words, the secondary system should behave rather like a human being might behave. Qualities such as competence, politeness and helpfulness are all more commonly associated with human-*human* interactions (or at least human-*robot* interactions; Goodrich and Olsen, 2003). Their correlation with naturalness suggests the car itself is in some ways being perceived as a semi-intelligent being, or possibly even anthropomorphically (as in Reeves and Nass, 1996)

Another higher-level pattern observed is that of **'clarity of design metaphor'** (Norman, 2013). Many of the 14 theme descriptions could be summarised by a phrase such as "It is clear what the car/designer is trying to do". Indeed, many participants had made comments similar to that phrase when interacting with competent natural-feeling systems. Clarity of metaphor and transparency of action and intent, appear to be important aspects of naturalness.

The final higher-level pattern observed is that many secondary car systems need to be designed for a user who is **carrying out a more important primary interaction concurrently** (Harvey and Stanton, 2013). This theme can arguably be observed in most of the 14 item dimensions above, especially Items 3 to 14. As suggested strongly by the participant observation study, secondary controls and systems are *not the main point of driving*, they are subsidiary and frequently operated only semi-consciously by a driver focusing their attention (visual or mental or both) on something more important - usually (but not always) the primary driving task. Numerous eyes-free or minimally-attended interactions with secondary controls were recorded. Yet the participant observation also showed many modern secondary systems appearing to demand drivers' *full* attention, especially those with screens and menus and detailed settings. This suggests manufacturers are not always designing secondary controls to be truly 'secondary'.

9.2 The Underlying Factors of Naturalness (RQ2)

Three factors were found to underlie the construct, according to principal components analysis. These may be described as (1) *Helpfulness and Control,* accounting for 27% of variance (2) *Strong Communication and Connection*, accounting for 21% of variance and (3) *Logical Location and Form*, accounting for 15% of variance. All had good to high reliability measured by Cronbach's alpha and the test as a whole had 'excellent' sampling adequacy. Each factor (or subscale) is represented by a minimum of three items in the normal naturalness measurement scale to increase reliability. This answers the second part of the research question. While Factors 2 and 3 have some parallels in established interaction naturalness, Factor 1 appears to be more novel and may be distinctive to automotive interactions.

9.3 The 15-Item Naturalness Measurement Scale (RQ3) and Comparison to Naturalness Interpretations in the Literature

In response to the third part of the research question, the results were used to develop the *first known measurement scale for driver-car naturalness* (and perhaps for any form of interaction naturalness). It gives an easily interpretable numerical score of likely perceived naturalness of car secondary systems. The measurement scale is easy to deploy, comprising just 15 bipolar items on five-point scales. In normal use it has been shown to take no more than five minutes per driver-system to administer when English is participants' first language. Using the scale with only native or highly fluent speakers is recommended given the nuances of some of the scale anchor wordings. The wording was derived from the language drivers themselves used when speaking freely at the very start of the research, and refined throughout. It uses non-specialist language which the data suggested is understandable and consistently interpreted by ordinary drivers. Each item is directly derived from one of the 14 high correlating dimensions in the final study's framework of naturalness (see Section 8.3 below) plus a 15th 'check' item which explicitly asks how 'natural/unnatural-feeling' the control is to use. This is the only item to use the word 'natural' in the scale. It is asked in the same way as all the other items and placed centrally, to avoid alerting drivers as to the true intention of the measurement scale, with the aim of avoiding prejudice (Steg et al, 2001). This 'check' item may be used to check face validity, reliability or fatigue effects.

The measurement tool was found to have acceptable content and criterion validity when 16 drivers' naturalness ratings of six secondary controls were compared to five expert opinions of the naturalness of the same systems in the same test car, there being no other ‘concurrent’ scale against which to validate it. A shorter ten-item heuristic version of the scale was produced and suggested but not yet validated. It is hoped this may be used as a ‘check list’ by automotive interface designers without customer involvement, because the literature review had suggested direct customer involvement may not always be feasible for early design ideation or revision of minor controls.

Name code: Age:		Car used:	0	1	2	3	4	System used:	N= /60
		Attribute ‘A’	Very A	Somewhat A	Neither A nor B	Somewhat B	Very B	Attribute ‘B’	x5/3= % cf Q10 %
E	*[Example question only] The control’s action felt:*	*hard*						*easy*	*Score*
1	Imagining the car is a person, the system seems:	Unhelpful						Helpful	
2	Imagining the car is a person, the system seems:	Rude						Polite	
3	The system seems:	Highly incompetent						Highly competent	
4	The car responds:	Unpredictably						Predictably	
5	When you do the action it feels like:	The car is fully in control						You are fully in control	
6	Operating the control feels:	Difficult						Easy	
7	Mentally the interaction is:	Highly demanding						Not at all demanding	
8	The interaction overall feels:	Counter intuitive						Intuitive	
								HAC=	*/32*
9	The communication between you and the car feels:	Artificial						Real	
10	The control’s response feels:	Delayed						Instant	
11	The car comes across as:	Uncommunicative						Informative	
12	Overall the interaction felt:	Unnatural						Natural *SCC=*	*/12*
								N=	*/4*
13	The control is located:	Illogically						Logically	
14	The shape and movement of the control:	Does not reflect its function at all						Closely reflects its function	
15	The input action required seems:	Completely unclear						Completely obvious	
								LFF=	*/12*

FIGURE 9.2 THE NATURALNESS MEASUREMENT SCALE

The discussion sections in most Chapters have considered at length how the detailed findings of each study related to interpretations of naturalness in the literature. Final consideration will now be given to how the final 15-Item naturalness measurement scale (reproduced as Figure 9.2 above) relates to existing interpretations of naturalness from the literature on general interaction naturalness:

- Richness of interaction (e.g. Jensen *et al.,* 2005) is perhaps suggested by items 9, 10 and 11;
- Physical/bodily interaction (e.g. Hornecker, 2011)) is perhaps suggested by items 9 and 10 but this is rather tenuous;
- Mimicry of some measurable physical property of the natural world (e.g. Goodman *et al.,* 2008) or natural physics (e.g. Malizia and Bellucci, 2012) or mimicry of some

natural human action on the world (e.g. Jacob *et al.,* 2008) is again only tenuously suggested, for example by item 9;

- Mimicry of human-human communication tendencies (e.g. Bickmore and Cassell, 2005, Marge et al 2010; Van Dam, 1997) is much more strongly implicated in the scale for example by items 1, 2, 3, 4, 9 and 11;
- Similarity to human-animal interaction (Flemish et al 2012) could again be related to items 1, 2, 3, 4, 9 and 11;
- Use of established 'cultural' HCI interactions (e.g. Malizia and Bellucci, 2012) could conceivably be related to items 4, 8, and 15;
- Interaction which closely matches human mental models (e.g. Goodrich and Olsen, 2003) might be linked to items 4, 6, 8, 14 and 15;
- Interaction with coherent design metaphor (Celentano and Dubois, 2014) is harder to judge because the metaphor differs from situation to situation and from person to person and often goes unstated, but it is feasible that items 4, 7, 8, 15 and 15 relate to metaphor.

What is generally missing from the established literature, compared to the findings above, are the 'humanlike', 'helpfulness' and 'politeness' aspects of naturalness suggested by this thesis's findings. Existing literature has also not explicitly suggested that issues of competence and control are related to naturalness, in contrast to this thesis's findings.

9.4 Conclusions Relating to Future or Highly Intelligent Cars (RQ4)

Some additional higher-level conclusions may be made about naturalness in intelligent and future secondary car control controls, which relate to the fourth part of the research question.

The naturalness of high levels of automation and fully self-driving cars requires further investigation. The single scale item that directly addresses automation (Item 5 'the car is fully in control – the driver is fully in control') appears to prejudice against it, because the anchor point 'the car feels fully in control' is designed to correlate with 'unnatural'. As discussed above, that correlation was derived from the typically software-driven 'piecemeal' automated systems tested in Chapter 7. Future cars are likely to feature self-driving and self-learning abilities. Poorly automated systems (for example with unconnected 'piecemeal' characteristics, or requiring constant supervision) would indeed probably be perceived as 'unnatural'. However, there was evidence that automated systems where the driver still perceives he/she has ultimate control, and which act competently, predictably, as a *single* intelligent being, and in a humanlike way, have a good chance of being perceived as *natural* by the scale. This was evidenced by numerous drivers reasoning out loud, when answering Item 5 about an apparently well-automated system, that although the car was technically in control, *they* were giving the command 'telling it what to do' – and therefore felt in control – giving a positive naturalness rating for this item. With respect to 'automated' driving, the more accurate description of 'self-driving' becomes pertinent. The most complete 'automated' cars will in fact use self-learning algorithms to 'self-drive', not pre-programmed software which needs to be constantly supervised or switched on and off. Fully 'self-driving' cars will effectively have a machine 'brain'. This arrangement has much more in common with the 'co-pilot' type systems described as 'natural' by drivers in the first two studies. Perhaps then, using a clear interaction design metaphor of 'driving co-pilot' rather than 'self-driving car' would create a more natural perception.

Free hand/arm gestures (i.e. 'air' gestures) do not appear to be perceived as a natural way of operating secondary systems at the present time. Communicating via contact gesture may be more natural but this is not yet common; no contact gesture interface was tested in this research other than three touchscreens with 'swipe' ability. Gestural controls (sometimes unfortunately termed 'natural user interfaces'; Norman, 2010) are currently the recipient of major OEM research and investment. This research suggests that where natural-*feeling* interfaces are the genuine aim, it may be more cost effective to invest in traditional principles of efficient human machine interaction. These include affordance of shape, input-response mapping, logical location, and multimodal feedback with built in redundancy. Once highly intelligent cars can demonstrate *natural language* understanding and communication, *vocal-auditory* communication may become the most natural way to interact with a car.

9.5 Conclusions about the Methodology Used

The use of qualitative human-centred research methods used and their adaptation to the car cabin may benefit future automotive researchers. In the first study the relatively simple decision to interview drivers *inside* their cars rather than in a laboratory took rather more time, but met many more criteria for ecological validity. The adaptation of Think Aloud and flexible modelling techniques to the automotive domain may hold potential for other researchers seeking broader perceptions and feelings related to car controls. More researchers should use these methods as a complement to traditional 'human performance' testing which tends to measure only how *competently* or how *quickly* users conduct a simulated task. 'Bench-testing' of unfamiliar automotive controls probably exaggerates their usability failings, but this may be desirable in some research. The 'breaching' method used in the exploratory workshops has not been recorded before in the automotive domain, yet as a means of ‘violating’ and exploring interaction ‘norms’ it provided many of the highest correlating dimensions, and probably led to the scale items being bipolar rather than unipolar. The ‘Wizard of Oz’ study involving a professional voice actor *and* researcher controlling an affordable but convincing prototype intelligent car on real urban roads may also be beneficial for future researchers’ studying future automotive interaction scenarios. This experimental set up used inexpensive cameras, smartphones and internet software. Prototypes designed to capture and respond vocally to drivers’ *voice* inputs, have potential to generate rich data and valuable insight into drivers ‘emotional states and perceptions. Drivers revealed far more *by voice* than was obtained by interpreting the subtle hand movements, face gestures or eye glances in the other observation. This resoundingly overcame the problem of interpreting ‘silent private’ interactions, and - as shown by the first study – a ‘speaking/listening car’ is entirely in line with drivers’ expectations for the future.

9.6 Recommendations for Future Work

This study, a first attempt at developing a rating scale for naturalness, had some limitations that future work might seek to overcome. Firstly, some research has suggested that perceptions of usability can be influenced by brand stereotypes and perceptions of quality (Wellings et al, 2008). In this research it is not known what contribution car brand stereotypes made, or indeed the potential effects of overfamiliarity with controls which drivers may use, misuse (or disuse) every day. A comparative study using blind testing inside cars that are unfamiliar to participants, or where the branding is effectively concealed (like in some consumer ‘car clinics’) may reduce these effects. Similarly, the validation took place in an expensive BMW car and it was not possible to hide the its brand. Further validation will be necessary on a larger sample of cars of different brands and market sectors, testing more controls and checking discriminant validity between *similar* systems.

With the measurement scale now in a compact form (just 15 items from the original 46) an attempt could be made to generalise the study to a larger more representative sample of the general car driving public than was possible in these exploratory studies. Traveling to meet participants in their cars was rather time consuming, with the researcher’s travel time far exceeding the time it took to administer the individual rating tests. The cabin of a car is also an intimate place where it is impossible for the researcher not to be intrusive. This may make it impossible for the drivers studied to be totally objective or to express themselves without inhibition. The naturalness measurement scale might therefore be administered *remotely* via email or a dedicated smartphone application, to a larger sample of the intended user population, or to more specific population samples (e.g. premium car drivers) in order to further evaluate correlations and reliability. This would permit a much larger sample. A safeguard could be built in to ensure participants *genuinely* do the rating exercise *inside* their cars – such as a forcing function that makes participants take a smartphone photograph of the secondary control in question before proceeding. It might also be administered remotely and on-road using platforms like the mobile Brunel Automotive Habitat Laboratory.

Further in the future, participant observations like those described in Chapter 6 could be conducted when it becomes possible to carry out affordable minimally intrusive *brain imaging* on drivers interacting with their cars' controls on road journeys. This could potentially provide direct scientific correlates of how drivers perceive and react to automotive interactions.

It was never presumed that naturalness would be a stable perception. Similar studies should also be conducted in five or ten years' time to check if there is a change in the meaning of naturalness at this time of rapid dashboard and smartphone evolution. Evolving technology *outside* the car may influence perceptions of what feels natural *inside* the car.

Finally, the 'Wizard of Oz' on-road simulation described in Chapter 6 might be extended (with dual controls and a real human 'co-pilot' driver concealed behind a curtain, for example) to explore what drivers *naturally* do while their car is driving itself. According to OEM concept films, drivers will hold business meetings, read books, socialise or even sleep during such journeys – but no ethnographic evidence apparently informs such concepts. Nor is it known, for example, whether drivers will instinctively face the direction of travel or feel comfortable facing backwards or sideways, nor whether they will be content with an entirely screen- or voice-based interface or instinctively demand some physical control to hold. Instinctive actions may be the most natural actions. What drivers instinctively do in a self-driving car will directly affect what secondary systems they will need in the future and how they might most naturally control them.

Simon Ramm, June 2017; January 2018

REFERENCES

Abbink, D. A., Mulder, M., & Boer, E. R. (2012). Haptic shared control: smoothly shifting control authority? *Cognition, Technology & Work*, *14*(1), 19-28.

Ahmadpour, N., Lindgaard, G., Robert, J.M. and Pownall, B., 2014. The thematic structure of passenger comfort experience and its relationship to the context features in the aircraft cabin. *Ergonomics*, *57*(6), pp.801-815.

Alvarez, I., Lopez-de Ipiña, K., Daily, S. B., & Gilbert, J. E. (2012) Emotional Adaptive Vehicle User Interfaces: moderating negative effects of failed technology interactions while driving. *Adjunct Proceedings of the 4th International Conference on Automotive User Interfaces* (AutomotiveUI '12), October 17–19, 2012, Portsmouth, NH.

Angulo, J., 2007. The Emotional Driver: A study of the driving experience and the road context.

Aslfallah, A. (2008). Design of automobile instrumentation.

Autocar Magazine, (2013) *A week in cars,* editorial on how power assistance has changed our view of 'natural' steering. 17th July 2013. Haymarket, London.

Baharin, H., Rintel, S., & Viller, S. (2013). Rhythms of the Domestic Soundscape: ethnomethodological soundwalks for phatic technology design. In *Human-Computer Interaction–INTERACT 2013* (pp. 463-470). Springer Berlin Heidelberg.

Bainbridge, L., 1983. Ironies of automation. *Automatica*, *19*(6), pp.775-779.

Baker, S. E., & Edwards, R. (2012). *How many qualitative interviews is enough.* Available at: http://eprints.ncrm.ac.uk/2273/4/how_many_interviews.pdf

Balkmar, D. (2012). *On Men and Cars: An Ethnographic Study of Gendered, Risky and Dangerous Relations* (Doctoral dissertation, Linköping).

Banks, V. A., Stanton, N. A., & Harvey, C. (2014a). Sub-systems on the road to vehicle automation: Hands and feet free but not 'mind'free driving. *Safety science*, *62*, 505-514.

Banks, V. A., Stanton, N. A., & Harvey, C. (2014b). What the drivers do and do not tell you: using verbal protocol analysis to investigate driver behaviour in emergency situations. *Ergonomics*, *57*(3), 332-342.

Barón, A., & Green, P. (2006). *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review* (No. UMTRI-2006-5). University of Michigan, Transportation Research Institute.

Baxter, A.L., Watcha, M.F., Baxter, W.V., Leong, T. and Wyatt, M.M., 2011. Development and validation of a pictorial nausea rating scale for children. *Pediatrics*, *127*(6), pp.e1542-e1549.

Bérard, F. and Rochet-Capellan, A., 2015, April. The transfer of learning as hci similarity: Towards an objective assessment of the sensory-motor basis of naturalness. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1315-1324). ACM.

Bernard, H. Russell (1994). *Research methods in anthropology: qualitative and quantitative approaches* (second edition). Walnut Creek, CA: AltaMira Press.

Beruscha, F., Wang, L., Augsburg, K., & Wandke, H. (2010, April). Do drivers steer toward or away from lateral directional vibrations at the steering wheel. In *Proc. 2nd European Conference on Human Centred Design for Intelligent Transport Systems* (pp. 227-236).

Beyer, H. and Holtzblatt, K., 1997. *Contextual design: defining customer-centered systems*. Elsevier.

Bhise, V.D., 2011. *Ergonomics in the automotive design process*. CRC Press.

Bickmore, T., & Cassell, J. (2005). Social Dialongue with Embodied Conversational Agents. In *Advances in natural multimodal dialogue systems* (pp. 23-54). Springer Netherlands.

Birrell, S. A., Young, M. S., & Weldon, A. M. (2010). Delivering smart driving feedback through a haptic pedal. In *Proceedings of the International Conference on Contemporary Ergonomics and Human Factors 2010* (pp. 431-439). Taylor & Francis Ltd

Bisping, R. (1997). Car interior sound quality: Experimental analysis by synthesis. *Acta Acustica united with Acustica*, *83*(5), 813-818.

Black, S. (1966). *Man and motor cars: An ergonomic study*. Secker & Warburg.

Bless, H., Fiedler, K., & Strack, F. (2004). *Social cognition: How individuals construct social reality*. Hove and New York: Psychology Press

Boess, S. U., & Kanis, H. (2007). Meaning in product use–A design perspective. *Product experience*, 305-332.

Bowler, M, 2012. HMI and Driver Distraction. Report by IQPC.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Brooke, J., 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), pp.4-7.

Brooke, J., 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), pp.4-7.

Brookhuis, K.A. and de Waard, D., 2010. Monitoring drivers’ mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention*, *42*(3), pp.898-903.

Broy, N., & Rümelin, S. (2012). Natural Visual User Interfaces–Beyond Input Modalities.

Broy, N., André, E., & Schmidt, A. (2012, October). Is stereoscopic 3D a better choice for information representation in the car?. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 93-100). ACM.

Burgess, M., King, N., Harris, M., & Lewis, E. (2013). Electric vehicle drivers’ reported interactions with the public: Driving stereotype change?. *Transportation research part F: traffic psychology and behaviour*, *17*, 33-44.

Burnett, G. E., & Mark Porter, J. (2001). Ubiquitous computing within cars: designing controls for non-visual use. *International Journal of Human-Computer Studies*, *55*(4), 521-531.

Burnett, G., & Irune, A. (2009, September). Drivers' quality ratings for switches in cars: assessing the role of the vision, hearing and touch senses. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 107-114). ACM.

Burns, P. C., Trbovich, P. L., Harbluk, J. L., & McCurdie, T. (2005, June). Evaluating one screen/one control multifunction devices in vehicles. In *The 19th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*.

Burr, V., 1998. Overview: Realism, relativism, social constructionism and discourse. *Social constructionism, discourse and realism*, pp.13-26.

Calefato, C., Montanari, R., & Tesauri, F. (2008). The Adaptive Automation Design. *New Developments*, 141

Carrabine, E., & Longhurst, B. (2008). Consuming the car: anticipation, use and meaning in contemporary youth culture. *The sociological review*, *50*(2), 181-196.

Cavanagh, S., 1997. Content analysis: concepts, methods and applications. *Nurse researcher*, *4*(3), pp.5-13.

Celentano, A. and Dubois, E., 2014. Metaphors, analogies, symbols: in search of naturalness in tangible user interfaces. *Procedia Computer Science*, *39*, pp.99-106.

Cellario, M. 2001, Human-centered intelligent vehicles: toward multimodal interface integration, IEEE Intelligent Systems, Vol. 16, No. 4, pp 78-81.

Chang, J. C., Lien, A., Lathrop, B., & Hees, H. (2009, September). Usability evaluation of a Volkswagen Group in-vehicle speech system. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 137-144). ACM.

Chang, W., Hwang, W., & Ji, Y. G. (2011). Haptic Seat Interfaces for Driver Information and Warning Systems. *International Journal of Human-Computer Interaction*, *27*(12), 1119-1132.

Charissis, V., Papanastasiou, S., Mackenzie, L., & Arafat, S. (2011). Evaluation of collision avoidance prototype head-up display interface for older drivers. In *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments* (pp. 367-375). Springer Berlin Heidelberg.

Chen, F., Wang, M., Duan, L (2015). Let Driver Know What is Going on in the Traffic *Letting Drivers Know What is Going on in Traffic-Chapter 10*. Department of Interaction Design and Technology Institute of Applied IT Chalmers University of technology, Sweden

Chen, H., Finin, T., & Joshi, A. (2003). An ontology for context-aware pervasive computing environments. *The Knowledge Engineering Review*, *18*(03), 197-207.

Cohen, P. R., & Oviatt, S. L. (1995). The role of voice input for human-machine communication. *proceedings of the National Academy of Sciences*, *92*(22), 9921-9927.

Consumer Reports Inc. (2014). *Consumer Reports 2014 Annual Auto Reliability Survey Finds Infotainment Systems are a Growing First-Year Reliability Plague*. [online] Available at: http://www.consumerreports.org/media-room/press-releases/2014/10/my-entry/ [Accessed 20/10/16].

Consumer Reports Inc. (2016). *Brand-by-Brand Guide to Car Infotainment Systems.* [online] Available at: http://www.consumerreports.org/cars-infotainment-system-brand-by-brand-guide/ [Accessed 07/01/17].

Coolican, H. (2009). *Research methods and statistics in psychology*. (5th Edition) Hodder Education, London.

Cooper, A., Reimann, R., & Cronin, D. (2012). *About face 3: The essentials of interaction design*. John Wiley & Sons.

Cornish, M., Cornish, D., & Dukette, D. (2009). *The essential 20: twenty components of an excellent health care team*. Dorrance Publishing.

Creswell, J. (2012). *Qualitative inquiry and research design: Choosing among five approaches*. Sage.

Cronbach, L.J., (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), pp.297-334.

Cycil, C., Perry, M., & Eardley, R. (2014, September). The HomeCar organiser: designing for blurring home-car boundaries. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 955-962). ACM.

Cycil, C.R., 2016. *Technology and the family car: situating media use in family life* (Doctoral dissertation, Brunel University London).

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993, February). Wizard of Oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces* (pp. 193-200). ACM.

Damiani, S., Deregibus, E. and Andreone, L., 2009. Driver-vehicle interfaces and interaction: where are they going?. *European transport research review*, *1*(2), pp.87-96.

Dant, T. (2004). The driver-car. *Theory, Culture & Society*, *21*(4-5), 61-79.

Dant, T., & Martin, P. J. (2001). By car: Carrying modern society. *Ordinary Consumption*, 143-157.

de Melo, G., Honold, F., Weber, M., Poguntke, M., & Berton, A. (2009, September). Towards a flexible UI model for automotive human-machine interaction. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 47-50). ACM.

Decrop, A. (1999). Triangulation in qualitative tourism research. *Tourism management*, *20*(1), 157-161.

Degen, M. (2014) Personal communication (email).

Demirbilek, O., & Sener, B. (2003). Product design, semantics and emotional response. *Ergonomics*, *46*(13-14), 1346-1360

deMunck, Victor C. & Sobo, Elisa J. (Eds) (1998). *Using methods in the field: a practical introduction and casebook*. Walnut Creek, CA: AltaMira Press.

Denzin, N.K. and Lincoln, Y.S., 1998. The Landscape of Qualitative Research: Theories and Issues.

Desmet, P. M., & Hekkert, P. (2002). The basis of product emotions. *Pleasure with products, beyond usability*, 60-68.

Desmet, P. M., & Hekkert, P. (2007). Framework of product experience. *International Journal of Design*, *1*(1), 57-66.

DeWalt, Kathleen M. & DeWalt, Billie R. (2002). *Participant observation: a guide for fieldworker*s. Walnut Creek, CA: AltaMira Press.

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J.D., Perez, M.A., Hankey, J., Ramsey, D.J., Gupta, S. and Bucher, C., 2006. *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment* (No. HS-810 593).

Dittmar, H. (1992). *The social psychology of material possessions: To have is to be*. Hemel Hempstead: Harvester Wheatsheaf.

Dogan, E., Steg, L., & Delhomme, P. (2011). The influence of multiple goals on driving behavior: The case of safety, time saving, and fuel saving. *Accident Analysis & Prevention*, *43*(5), 1635-1643.

Doshi, A., Cheng, S. Y., & Trivedi, M. M. (2009). A novel active heads-up display for driver assistance. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *39*(1), 85-9 3.

Dourish, P., 2004. *Where the action is: the foundations of embodied interaction*. MIT press.

Drummond, G. and Ensor, J., 2006. *Introduction to marketing concepts*. Routledge.

Dunne, A. (1999). Hertzian tales. *Electronic products, aesthetic experience and critical design*

Dvorsky, G. (2013). http://io9.com/5974468/the-most-common-cognitive-biases-that-prevent-you-from-being-rational

Eckoldt, K., Knobel, M., Hassenzahl, M., & Schumann, J. (2012). An Experiential Perspective on Advanced Driver Assistance Systems. *it-Information Technology*, *54*(4), 165-171.

Ekkekakis, P. (2013). *The measurement of affect, mood, and emotion: A guide for health-behavioral research*. Cambridge University Press.

Ekman, E.P. (1970) *Universal Facial Expressions of Emotion*. California Mental Health Research Digest, Vol. 8.

Ekman, E.P and Friesen, W. V. (1978) *Facial Action Coding System Investigator's Guide*, Consulting Psychologist Press. Palo Alto, CA

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *J.Advanced Nursing*, *62*(1), 107-115.

Ericsson, K. A. Simon. HA (1984). *Protocol analysis: Verbal reports as data*.

Eyben, F., Wöllmer, M., Poitschke, T., Schuller, B., Blaschke, C., Färber, B., & Nguyen-Thien, N. (2010). Emotion on the road—necessity, acceptance, and feasibility of affective computing in the car. *Advances in Human-Computer Interaction*, *2010*.

Farhadi-Niaki, F., Gerroir, J., Arya, A., Etemad, S. A., Laganière, R., Payeur, P., & Biddle, R. (2013, February). Usability Study of Static/Dynamic Gestures and Haptic Input as Interfaces to 3D Games. In *ACHI 2013, The Sixth International Conference on Advances in Computer-Human Interactions* (pp. 315-323).

Fernandez-Martinez et al. (2012) Interacting with Computers

Feuš, M. (2013). *The Haptic Drive: Haptic feedback for car interfaces* (Doctoral dissertation, Umeå University).

Field, A., 2009. *Discovering statistics using SPSS*. Sage publications.

Finstad, K., 2010. Response interpolation and scale sensitivity: Evidence against 5-point scales. *Journal of Usability Studies*, *5*(3), pp.104-110.

Fischer, R.A., 1984. Testing written communicative competence in French. *The Modern Language Journal*, *68*(1), pp.13-20.

Fitch, G. M., Hankey, J. M., Kleiner, B. M., & Dingus, T. A. (2011). Driver comprehension of multiple haptic seat alerts intended for use in an integrated collision avoidance system. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(4), 278-290.

Fleischmann, T., 2007, July. Model based HMI specification in an automotive context. In *Symposium on Human Interface and the Management of Information* (pp. 31-39). Springer Berlin Heidelberg.

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., & Beller, J. (2012). Towards a dynamic balance between humans and authority, ability, responsibility and control in shared and cooperative control situations. *Cognition, Technology & Work*, 1-16.

Flemisch, F., Nashashibi, F., Rauch, N., Schieben, A., Glaser, S., Temme, G., & Kaussner, A. (2010, June). Towards Highly Automated Driving: Intermediate report on the HAVEit-Joint System. In *3rd European Road Transport Research Arena, TRA 2010*

Flemisch, F.O., Adams, C.A., Conway, S.R., Goodrich, K.H., Palmer, M.T. and Schutte, P.C., 2003. The H-Metaphor as a guideline for vehicle automation and interaction.

Fowler, S. (2015) CES 2015: Mercedes F 015 shows its vision of the future. Auto Express, 6 Jan: http://www.autoexpress.co.uk/mercedes/89925/ces-2015-mercedes-f-015-shows-its-vision-of-the-future

Fraine, G., Smith, S. G., Zinkiewicz, L., Chapman, R., & Sheehan, M. (2007). At home on the road? Can drivers' relationships with their cars be associated with territoriality?. *Journal of environmental psychology*, *27*(3), 204-214.

Friese, S. (2014). *Qualitative data analysis with ATLAS. ti*. Sage.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall, 1967, p. 32

Gaspar, J., Fontul, M., Henriques, E. and Silva, A., 2014. User satisfaction modeling framework for automotive audio interfaces. *International Journal of Industrial Ergonomics*, *44*(5), pp.662-674.

Gaver, B., Dunne, T., & Pacenti, E. (1999). Design: cultural probes. *interactions*, *6*(1), 21-29.

Geiser, G., 1985. Man machine interaction in vehicles. *ATZ*, *87*(74-77), p.56.

Gellatly, A. W., Hansen, C., Highstrom, M., & Weiss, J. P. (2010, November). Journey: General Motors' move to incorporate contextual design into its next generation of automotive HMI designs. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 156-161). ACM.

Gentner, D. and Grudin, J., 1985. The evolution of mental metaphors in psychology: A 90-year retrospective. *American Psychologist*, *40*(2), p.181.

Giacomin, J. (2012a) Human Centred Design of 21st Century Automobiles, Report by Human Centred Design Institute, Brunel University Uxbridge.

Giacomin, J. (2012b) What is Human Centred Design? Paper presented at 10º Congresso Brasileiro de Pesquisa e Desenvolvimento em Design, São Luís (MA)

Giacomin, J. (2014). What Is Human Centred Design?. *The Design Journal*,*17*(4), 606-623.

Giacomin, J., Ramm, S. (2013) There's more to safe driving than information and decisions. Fleet Safety Conference & Awards 2013, 13-14 June, St John's Hotel, Solihull, UK. *Report by Human Centred Design Institute, Brunel University Uxbridge*, Middlesex.

Giboreau, A., Navarro, S., Faye, P., & Dumortier, J. (2001). Sensory evaluation of automotive fabrics: the contribution of categorization tasks and non verbal information to set-up a descriptive method of tactile properties. *Food quality and preference*, *12*(5), 311-322.

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. *The American journal of psychology*, *51*(3), 453-471.

Gkouskos, D., Normark, C.J. and Lundgren, S., 2014. What drivers really want: Investigating dimensions in automobile user needs. *International Journal of Design*, *8*(1).

Glaser, B., & Strauss, A. (2009). *The discovery of grounded theory: Strategies for qualitative research*. Transaction Books.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report*, *8*(4), 597-606.

Gold, Raymond L. (1958). Roles in sociological field observations. *Social Forces, 36*, 217-223.

Gomez, R.E., Popovic, V. and Bucolo, S., 2004. Driving: The emotional experience and automotive design.

Goodman, E., Kuniavsky, M., & Moed, A. (2013). Observing the User Experience: A Practitioner's Guide to User Research. *Professional Communication, IEEE Transactions on*, *56*(3), 260-261.

Goodman, Teresa, Ruth Montgomery, A. Bialek, Alistair Forbes, M. Rides, A. Whitaker, K. Overliet, F. McGlone, and G. W. A. M. Heijden. "Measurement of Naturalness." *Measurement* (2008).

Goodrich, M.A. and Olsen, D.R., 2003, October. Seven principles of efficient human robot interaction. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on* (Vol. 4, pp. 3942-3948). IEEE.

Graf, S., Spiessl, W., Schmidt, A., Winter, A., & Rigoll, G. (2008, April). In-car interaction using search-based user interfaces. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1685-1688). ACM.

Graham, R. (1999). Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics*, *42*(9), 1233-1248.

Graneheim, U., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse education today*, *24*(2), 105-112.

Green, P. (1999, September). Estimating compliance with the 15-second rule for driver-interface usability and safety. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 43, No. 18, pp. 987-991). SAGE Publications.

Green, P. (2012) In-vehicle UI standards, Paul Green, University of Michigan. Tutorial at: 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive UI 2012) Portsmouth, New Hampshire, USA,

Groeger, J. A. (2002). Trafficking in cognition: applying cognitive psychology to driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(4), 235-248.

Gross, R. (2012). *Psychology: The Science of Mind and Behaviour 6th Edition*. Hachette UK.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field methods*, *18*(1), 59-82.

Guilford, T. and Dawkins, M.S., 1991. Receiver psychology and the evolution of animal signals. *Animal behaviour*, *42*(1), pp.1-14.

Gunn, W. and Donovan, J., 2012. Design anthropology: an introduction. *Design and anthropology*, *5*, pp.1-16.

Guo, F., & Fang, Y. (2013). Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*, *61*, 3-9.

Hagman, Olle (2010) Driving Pleasure: A Key Concept in Swedish Car Culture, *Mobilities* 5(1), pp. 25-39.

Hajek, W., Gaponova, I., Fleischer, K.H. and Krems, J., 2013. Workload-adaptive cruise control–A new generation of advanced driver assistance systems. *Transportation research part F: traffic psychology and behaviour*, *20*, pp.108-120.

Hancock, P. A. (Ed.). (1999). *Human Performance and Ergonomics: Perceptual and Cognitive Principles*. Access Online via Elsevier.

Hanington, B. and Martin, B., 2012. *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Publishers.

Harper Collins, (2017). *Definitions of 'natural'* [online]. Available at: https://www.collinsdictionary.com/dictionary/english [Accessed 28/03/14]

Harris, D., Chan-Pensley, J. and McGarry, S., 2005. The development of a multidimensional scale to evaluate motor vehicle dynamic qualities. *Ergonomics*, *48*(8), pp.964-982.

Harris, H., & Nass, C. (2011, May). Emotion regulation for frustrating driving contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 749-752). ACM.

Hart, S.G. and Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, pp.139-183.

Harvey, C. and Stanton, N.A., 2013. *Usability evaluation for in-vehicle systems*. Crc Press.

Heim, S. (2008). *The resonant interface*. Pearson Education.

Helander, M.G. ed., 2014. *Handbook of human-computer interaction*. Elsevier.

Henderson, K. A. (1991). Dimensions of choice: A qualitative approach to recreation, parks, and leisure research. State College: Venture Publishing.

Herd, A. P. (2013). R2DFord: Autonomous Vehicles and the Legal Implications of Varying Liability Structures.

Hill, D. (2008). *Emotionomics*. Leveraging emotions for business, Revised edition

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklade, A.L. and Christ, R.E., 1992. Comparison of four subjective workload rating scales. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(4), pp.429-439.

Hinkin, T.R., 1998. A brief tutorial on the development of measures for use in survey questionnaires. *Organisational research methods*, *1*(1), pp.104-121.

Hinkin, T.R., Tracey, J.B. and Enz, C.A., 1997. Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research*, *21*(1), pp.100-120.

Hinkin, T.R., Tracey, J.B. and Enz, C.A., 1997. Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research*, *21*(1), pp.100-120.

Hintikka, J., 1979. Language-games. In *Game-Theoretical Semantics* (pp. 1-26). Springer Netherlands.

Ho, C., & Spence, C. (2012). *The multisensory driver: Implications for ergonomic car interface design*. Ashgate Publishing.

Ho, C., Reed, N., & Spence, C. (2006). Assessing the effectiveness of" intuitive" vibrotactile warning signals in preventing front-to-rear-end collisions in a driving simulator. *Accident analysis and prevention*, *38*(5), 988-996.

Hoffman, G., Gal-Oz, A., David, S., & Zuckerman, O. (2013, June). In-car game design for children: child vs. parent perspective. In *Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 112-119). ACM.

Hornecker, E. (2011). The role of physicality in tangible and embodied interactions. *interactions*, *18*(2), 19-23.

Hung, V. C., & Gonzalez, A. J. (2013). Context‐Centric Speech‐Based Human–Computer Interaction. *International Journal of Intelligent Systems*.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. *Handbook of cognitive task design*, *8*, 147-169.

Interregs Ltd. (2012). http://www.interregs.com/spotlight.php?id=114 EU Regulation on Gear Shift Indicators Published EU Regulation on Gear Shift Indicators Published

Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., & Zigelbaum, J. (2008, April). Reality-based interaction: a framework for post-WIMP interfaces. In *CHI-CONFERENCE-* (Vol. 1, p. 201). ASSOCIATION FOR COMPUTING MACHINERY INC.

Jaspers, M. W. (2006). A comparison of usability methods for testing interactive health technologies: methodological aspcts and empirical evidence. *International journal of medical informatics*, *78*(5), 340-353.

Jaspers, M.W.M., 2008. The Think Aloud Method and User Interface Design. *End-User Computing: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*.

Jensen, M. (1999). Passion and heart in transport—a sociological analysis on transport behaviour. *Transport Policy*, *6*(1), 19-33.

Jensen, M. V., Buur, J., & Djajadiningrat, T. (2005, August). Designing the user actions in tangible interaction. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility* (pp. 9-18). ACM.

Jindo, T., & Hirasago, K. (1997). Application studies to car interior of Kansei engineering. *International Journal of Industrial Ergonomics*, *19*(2), 105-114.

Johnson, B. D. (2011). Science fiction prototyping: designing the future with science fiction. *Synthesis Lectures on Computer Science*, *3*(1), 1-190.

Jordan, P. W. (2002). *Designing pleasurable products: An introduction to the new human factors*. CRC press.

Jordan, P.W., 1998. Human factors for pleasure in product use. *Applied ergonomics*, *29*(1), pp.25-33.

Juhlin, O. (2011). Social Media on the Road: Mobile Technologies and Future Traffic Research. *Multimedia, IEEE*, *18*(1), 8-10.

Karlsson, B. S., Aronsson, N., & Svensson, K. A. (2003). Using semantic environment description as a tool to evaluate car interiors. *Ergonomics*, *46*(13-14), 1408-1422.

Kawulich, B. (2005). Participant Observation as a Data Collection Method. ***Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 6***(2). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/466/996

Kazi, T. A., Stanton, N. A., Walker, G. H., & Young, M. S. (2007). Designer driving: drivers' conceptual models and level of trust in adaptive cruise control. *International Journal of Vehicle Design*, *45*(3), 339-360.

Kern, D., 2012. *Supporting the Development Process of Multimodal and Natural Automotive User Interfaces* (Doctoral dissertation, Universität Duisburg-Essen, Fakultät für Wirtschaftswissenschaften» Fachgebiet Wirtschaftsinformatik).

Kern, D., & Pfleging, B. (2013). Supporting interaction through haptic feedback in automotive user interfaces. *interactions*, *20*(2), 16-21.

Kern, D., & Schmidt, A. (2009, September). Design space for driver-based automotive user interfaces. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 3-10). ACM.

Kienle, M., Damböck, D., Kelsch, J., Flemisch, F., & Bengler, K. (2009, September). Towards an H-Mode for highly automated vehicles: driving with side sticks. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 19-23). ACM

Knapper, C. K., & Cropley, A. J. (1980). Interpersonal factors in driving. *Applied Psychology*, *29*(4), 415-438.

Koshar, R. (2001). On the History of the Automobile in Everyday Life. *Contemporary European History*, *10*(1), 143-154.

KPMG LLP, 2015, Connected and Autonomous Vehicles: The UK Economic Opportunity. Report produced for Society of Motor Manufacturers and Traders (SMMT), March.

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *The American journal of occupational therapy*, *45*(3), 214-222.

Krippendorff, K., 2004. *Content analysis: An introduction to its methodology*. Sage.

Kristensson, P., Matthing, J., & Johansson, N. (2008). Key strategies for the successful involvement of customers in the co-creation of new technology-based services. *International Journal of Service Industry Management*, *19*(4), 474-491.

Kumar, M., & Kim, T. (2005, April). Dynamic speedometer: dashboard redesign to discourage drivers from speeding. In *CHI'05 extended abstracts on Human factors in computing systems* (pp. 1573-1576). ACM.

Kun, A. L., Boyle, L. N., Reimer, B., & Riener, A. (2013). AutomotiveUI: Interacting with Technology in Vehicles. *Pervasive Computing, IEEE*, *12*(2), 80-82.

Lakoff, G. and Johnson, M., 2008. *Metaphors we live by*. University of Chicago press.

Lanks, B. (2015) http://www.bloomberg.com /news/articles/2015-01-08/ what-drivers-want-design-lessons-from-fords-mytouch-debacle. Accessed on 6th April 2015.

Larsson, S., & Villing, J. (2007). The dico project: A multimodal menu-based in-vehicle dialogue system. In *Proceedings of the 7th International Workshop on Computational Semantics (IWCS-7)* (p. 4).

Laurier, E. (2002). The region as a socio-technical accomplishment of mobile workers. *Wireless World: Social and Interactional Aspects of the Mobile Age, Computer Supported Cooperative Work*, 46-61.

Laurier, E. (2005). Searching for a parking space.

Laurier, E., & Philo, C. (1998). Meet You At Junction 17: A socio-technical and spatial study of the mobile office. *Glasgow: Dept. of Geography, University of Glasgow & ESRC, Swindon*.

Lavring, D. (2012) *Study Shows In-Car Technology Is a Pain-Point For Consumers*. [online] Available at: https://www.wired.com/2012/06/jd-power-2012-iqs/ [Accessed 22/10/16].

Lawshe, C.H. (1975). A quantitative approach to content validity. Personnel Psychology, 28, 563–575. doi:10.1111/j.1744-6570.1975.tb01393.x

Lee, S.K., 2008. Objective evaluation of interior sound quality in passenger cars during acceleration. *Journal of Sound and Vibration*, *310*(1), pp.149-168.

Liu, Y. C., & Wen, M. H. (2004). Comparison of head-up display (HUD) vs. head-down display (HDD): driving performance of commercial vehicle operators in Taiwan. *International Journal of Human-Computer Studies*, *61*(5), 679-697

Lo, V. E. W., & Green, P. A. (2012). Development and Evaluation of Automotive Speech Interfaces: Useful Information from the Human Factors and Related Literature.

Loasby, M., 1995, Is refinement and ICE eroding good handling? Automotive Engineer, 20, 2 – 3. 249 National Academies Press (2012) http://www.nap.edu/openbook.php?record_id=2308&page=16 accessed on 6/12/12

Loehmann, S., 2012. Sneaking Interaction Techniques into Electric Vehicles. *Adj. Proc. AutoUI 2012*, pp.111-112.

Loukopoulos, P., & Scholz, R. W. (2004). Sustainable future urban mobility: using 'area development negotiations' for scenario assessment and participatory strategic planning. *Environment and Planning A*, *36*(12), 2203-2226.

Makri, S., Blandford, A., & Cox, A. L. (2011). This is what I'm doing and why: Methodological reflections on a naturalistic think-aloud study of interactive information behaviour. *Information Processing & Management*, *47*(3), 336-348.

Malizia, A., & Bellucci, A. (2012). The artificiality of natural user interfaces. *Communications of the ACM*, *55*(3), 36-38.

Marge, M., Miranda, J., Black, A. W., & Rudnicky, A. I. (2010, September). Towards improving the naturalness of social conversations with dialogue systems. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 91-94). Association for Computational Linguistics.

Marinov, B., Mandadjieva, S. and Kostianev, S., 2008. Pictorial and verbal category-ratio scales for effort estimation in children. *Child: care, health and development*, *34*(1), pp.35-43.

Marsh, P. E., & Collett, P. (1986). Driving passion: The psychology of the car.

Marshall, Catherine & Rossman, Gretchen B. (1989). *Designing qualitative research*. Newbury Park, CA: Sage.

Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 11, No. 3, August).

McFarland, D., 1981. *Oxford companion to animal behavior*. Oxford University Press.

McDowell, I., 2006. *Measuring health: a guide to rating scales and questionnaires*. Oxford university press.

McGilchrist, I. (2009). *The master and his emissary: The divided brain and the making of the western world*. Yale University Press.

McGwin G., & Brown, D. (1999). Characteristics of traffic crashes among young, middle-aged, and older drivers. *Accident Analysis & Prevention*, *31*(3), 181-198.

McLeod, S. A. (2013). What is Validity? Retrieved from www.simplypsychology.org/validity.html Accessed on 22/7/15.

Meschtscherjakov, A., Tscheligi, M., Szostak, D., Krome, S., Pfleging, B., Ratan, R., ... & Ju, W. (2016, May). HCI and Autonomous Vehicles: Contextual Experience Informs Design. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 3542-3549). ACM.

Meschtscherjakov, A., Tscheligi, M., Szostak, D., Ratan, R., McCall, R., Politis, I. and Krome, S., 2015, April. Experiencing autonomous vehicles: Crossing the boundaries between a drive and a ride. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2413-2416). ACM.

Meschtscherjakov, A., Wilfinger, D., Gridling, N., Neureiter, K., & Tscheligi, M. (2011, November). Capture the car!: qualitative in-situ methods to grasp the automotive context. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 105-112). ACM.

Mesken, J., Hagenzieker, M.P., Rothengatter, T. and de Waard, D., 2007. Frequency, determinants, and consequences of different drivers' emotions: An on-the-road study using self-reports,(observed) behaviour, and physiology. *Transportation research part F: traffic psychology and behaviour*, *10*(6), pp.458-475.

Messick, S., 1989. Validity. In RL Linn (Ed.), *Educational measurement* (pp. 13-103). Washington, DC: American Council on Education and National Council on Measurement in Education.

Miles, M.B. and Huberman, A.M., 1985. *Qualitative data analysis*. Newbury Park,, CA: Sage.

Miller, D. (2001). *Car cultures*. Berg Publishers.

Möller, S. and Raake, A., 2014. Quality of Experience. *New York, US: Springer*.

Morse, J. M., & Field, P. A. (1995). *Qualitative research methods for health professionals*. Sage.

Morse, J.M., 1991. Approaches to qualitative-quantitative methodological triangulation. *Nursing research*, *40*(2), pp.120-123.

Mruck, K. and Breuer, F., 2003, May. Subjectivity and reflexivity in qualitative research—The FQS issues. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 4, No. 2).

Nasoz, F., Lisetti, C. L., & Vasilakos, A. V. (2010). Affectively intelligent and adaptive car interfaces. *Information Sciences*, *180*(20), 3817-3836.

Nass, C., Jonsson, I. M., Harris, H., Reaves, B., Endo, J., Brave, S., & Takayama, L. (2005, April). Improving automotive safety by pairing driver emotion and car voice emotion. In *CHI'05 extended abstracts on Human factors in computing systems* (pp. 1973-1976). ACM.

Neale, D. C., & Carroll, J. M. (1997). The role of metaphors in user interface design. *Handbook of human-computer interaction*, *2*, 441-462.

Neureiter, K., Meschtscherjakov, A., Wilfinger, D. and Tscheligi, M., 2011. Investigating the usage of multifunctional rotary knobs in the center stack with a contextual inquiry. *Proceedings of EA AUI 2011*.

NHTSA, 2012 Traffic Safety Facts 2012: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. DOT HS 812 032. *National Highway Traffic Safety Administration National Center for Statistics and Analysis U.S. Department of Transportation.* Washington

Norman, D. (2004). Emotional design. *Basic Books, New York*.

Norman, D. A. (1990). The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *327*(1241), 585-593.

Norman, D. A. (2005). There's an automobile in HCI's future, interactions, v. 12 n. 6. *November+ December*.

Norman, D. A. (2010). Natural user interfaces are not natural. *interactions*, *17*(3), 6-10.

Norman, D.A., 2013. *The design of everyday things: Revised and expanded edition*. Basic books.

Noy, Y. I. (Ed.). (1997). *Ergonomics and safety of intelligent driver interfaces*. Routledge.

O'hara, K., Harper, R., Mentis, H., Sellen, A., & Taylor, A. (2013). On the naturalness of touchless: putting the “interaction” back into NUI. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *20*(1), 5.

Ohn-Bar, E., Tran, C., & Trivedi, M. (2012, October). Hand gesture-based visual user interface for infotainment. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 111-115). ACM.

Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010, April). Think-aloud protocols: a comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2381-2390). ACM.

Osgood, C. E. (1957). *The measurement of meaning* (Vol. 47). University of Illinois Press.

Patton, M.Q., 1990. *Qualitative evaluation and research methods* . SAGE Publications, inc.

Perterer, N., Sundström, P., Meschtscherjakov, A., Wilfinger, D., & Tscheligi, M. (2013, February). Come drive with me: an ethnographic study of driver-passenger pairs to inform future in-car assistance. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 1539-1548). ACM.

Petersson, L., Fletcher, L., & Zelinsky, A. (2005, September). A framework for driver-in-the-loop driver assistance systems. In *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE* (pp. 771-776). IEEE.

Pieraccini, R., & Huerta, J. (2005). Where do we go from here? Research and commercial spoken dialog systems. In *6th SIGdial Workshop on Discourse and Dialogue*.

Pitts, M. J., Williams, M. A., Wellings, T., & Attridge, A. (2009, September). Assessing subjective response to haptic feedback in automotive touchscreens. In *Proceedings of the 1st international Conference on Automotive User interfaces and interactive Vehicular Applications* (pp. 11-18). ACM.

Pressnell, J. (1994). *The Mini* (No. 299). Osprey Publishing.

Rakotonirainy, A., & Steinhardt, D. (2009, September). In-vehicle technology functional requirements for older drivers. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 27-33). ACM.

Ramm, S., Giacomin, J., Robertson, D., & Malizia, A. (2014, September). A First Approach to Understanding and Measuring Naturalness in Driver-Car Interaction. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-10). ACM.

Ranney, T. (1994). Models of driving behavior: a review of their evolution. *Accident Analysis & Prevention*, *26*(6), 733-750.

Rasmussen, J., 1983. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics*, (3), pp.257-266.

Redshaw, Sarah (2008) *In the Company of Cars: Driving as a Social and Cultural Practice*. Aldershot, England: Ashgate Pub.

Reeves, B., & Nass, C. (1996). *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge university press.

Reich, D. and Stark, R., 2016. Head-Mounted Displays for HCI Validations While Driving. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *10*(10), pp.1765-1769.

Riener, A. (2012, October). Driver-vehicle confluence or how to control your car in future?. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 217-224). ACM.

Riener, A. (2012). Gestural Interaction in Vehicular Applications. *Computer*, *45*(4), 42-47.

Rigas, G., Goletsis, Y., Bougia, P., & Fotiadis, D. I. (2011). Towards driver's state recognition on real driving conditions. *International Journal of Vehicular Technology*, *2011*.

Risser, R., Lexell, E.M., Bell, D., Iwarsson, S. and Ståhl, A., 2015. Use of local public transport among people with cognitive impairments–A literature review. *Transportation research part F: traffic psychology and behaviour*, *29*, pp.83-97.

Robinson, J.P., Shaver, P.R. and Wrightsman, L.S., 1991. Criteria for scale selection and evaluation. *Measures of personality and social psychological attitudes*, *1*(3), pp.1-16.

Rowson, J. and Young, J. 2011 (December. Cabbies, Costs and Climate Change Cabbies, Costs and Climate Change:An engaged approach to fuel efficient behaviour. *RSA Reports.* Royal Society of the Arts, London.

Saffarian, M., de Winter, J. C. F., & Happee, R. (2012, September). Automated Driving: Human-factors issues and design solutions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 2296-2300). SAGE Publications.

Saldaña, J., 2015. *The coding manual for qualitative researchers*. Sage.

Sarter, N.B. and Woods, D.D., 1995. How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), pp.5-19.

Sauro, J. (2010) Can you use the SUS for websites? http://www.measuringu.com/blog/sus-websites.php Accessed om 5/11/16

Schieben, A., Griesche, S., Hesse, T., Fricke, N., & Baumann, M. (2014). Evaluation of three different interaction designs for an automatic steering intervention. *Transportation research part F: traffic psychology and behaviour*, *27*, 238-251.

Schifferstein, H. N., & Hekkert, P. (Eds.). (2011). *Product experience*. Access Online via Elsevier.

Schmidt, A. (2012) Tutorial Introduction to Automotive User Interfaces, University of Stuttgart, Germany accessed at http://auto-ui.org 19-10-2012

Schmidt, A., Dey, A. K., Kun, A. L., & Spiessl, W. (2010, April). Automotive user interfaces: human computer interaction in the car. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems* (pp. 3177-3180). ACM.

Schmidt, A., Spiessl, W., & Kern, D. (2010). Driving automotive user interface research. *Pervasive Computing, IEEE*, *9*(1), 85-88.

Scjmidt-Nielsen, B., Harsham, B., Raj, B., Forlines, C. (2008) *Handbook of research on user interface design and evaluation for mobile technology*. Information Science Reference, 2008.

Seppelt, B. D., & Lee, J. D. (2007). Making adaptive cruise control (ACC) limits visible. *International journal of human-computer studies*, *65*(3), 192-205.

Shedroff, N., & Noessel, C. (2012). *Make it So: Interaction Design Lessons from Science Fiction*. O'Reilly Media, Inc..

Sheik-Nainar, M., Huber, J., Bose, R. and Matic, N., 2016, May. Force-enabled TouchPad in Cars: Improving Target Selection using Absolute Input. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2697-2704). ACM.

Sheller, M. (2004). Automotive emotions feeling the car. *Theory, culture & society*, *21*(4-5), 221-242.

Sheller, M., & Urry, J. (2002). The city and the car. *International Journal of Urban and Regional Research*, *24*(4), 737-757.

Sheridan, T. B. (2002). *Humans and automation: System design and research issues*. John Wiley & Sons, Inc..

Shi, C., Lee, H.J., Kurczal, J. and Lee, A., 2012, October. Routine driving infotainment app: Gamification of performance driving. In *Adjunct Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 181-183).

Shimai, S., Tanaka, M., & Terasaki, M. (1990). Pleasantness and unpleasantness of environmental sounds. *Perceptual and motor skills*, *71*(3), 960-962.

Sivak, M. (1996). The information that drivers use: is it indeed 90% visual? *PERCEPTION-LONDON-*, *25*, 1081-1090.

Smith, J., Bekker, H., & Cheater, F. (2011). Theoretical versus pragmatic design in qualitative research. *Nurse researcher*, *18*(2), 39-51.

SMMT (2016). *SMMT Motor Industry Facts 2016* [online]. Report by Society of Motor Manufacturers and Traders. Available at: https://www.smmt.co.uk/wp-content/uploads/sites/2/SMMT-Motor-Industry-Facts-2016_v2-1.pdf [Accessed on 23/12/16].

Spies, R., Ablaßmeier, M., Bu bb, H. and Hamberger, W. (2009). Augmented interaction and visualisation in the automotive domain. In *International Conference on Human-Computer Interaction* (pp. 211-220). Springer Berlin Heidelberg.

Spradley, J.P., 1979. The ethnographic interview.

Spradley, James P. (1980). *Participant observation*. New York: Holt, Rinehart and Winston.

Stanton, N. A., & Young, M. S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics*, *48*(10), 1294-1313.

Steg, L. (2003). Can public transport compete with the private car?. *IATSS RESEARCH*, *27*(2), 27.

Steg, L. (2005). Car use: lust and must. Instrumental, symbolic and affective motives for car use. *Transportation Research Part A: Policy and Practice*, *39*(2), 147-162.

Steg, L., Vlek, C., & Slotegraaf, G. (2001). Instrumental-reasoned and symbolic-affective motives for using a motor car. *Transportation Research Part F: Traffic Psychology and Behaviour*, *4*(3), 151-169.

Stevens, S.S., 1946. On the theory of scales of measurement.

Stewart, D. W., & Shamdasani, P. N. (2014). *Focus groups: Theory and practice* (Vol. 20). Sage Publications.

Strauss, A. and Corbin, J., 1990. *Basics of qualitative research* (Vol. 15). Newbury Park, CA: Sage.

Streeck, J., Goodwin, C., & LeBaron, C. (Eds.). (2011). *Embodied interaction: Language and body in the material world* (Vol. 8). Cambridge University Press.

Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.

Susini, P., Misdariis, N., Lemaitre, G. and Houix, O., 2012. Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback. *Journal on Multimodal User Interfaces*, *5*(3-4), pp.175-186.

Takeuchi, S., Nishikawa, M., Suzuki, T., Shinzato, T., & Miyata, M. (2000). Ergonomic considerations in steering wheel controls. In *Sae Publication Sp-1539. Human Factors In 2000: Driving, Lighting, Seating Comfort, And Harmony In Vehicle Systems:* Papers Presented At The Sae 2000 World Congress, March 6-9, 2000, Detroit, Michigan, USA (Sae Technical Paper 2000-01-0169).

Tal, E., 2013. Old and new problems in philosophy of measurement. *Philosophy Compass*, *8*(12), pp.1159-1173.

Tal, Eran, "Measurement in Science", *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2015/entries/measurement-science/>.

Tertoolen, G., van Kreveld, D., & Verstraten, B. (1998). Psychological resistance against attempts to reduce private car use. *Transportation Research Part A: Policy and Practice*, *32*(3), 171-181.

Tonetto, L. M., & Desmet, P. M. (2016). Why we love or hate our cars: A qualitative approach to the development of a quantitative user experience survey. *Applied ergonomics*, *56*, 68-74.

Tractinsky, N., Abdu, R., Forlizzi, J., & Seder, T. Evaluating Design Language for Vehicle Instrument Clusters. *AutomotiveUI 2009*, 9.

Tscheligi, M. (2012). User Experience Design for Vehicles. Christian doppler laboratory: contextual interfaces. *Tutorial for conference AUI 2012*

Urry, J. (2006). Inhabiting the car. *The Sociological Review*, *54*, 17-31.

Vadeby, A., Wiklund, M., & Forward, S. (2011). Car drivers' perceptions of electronic stability control (ESC) systems. *Accident Analysis & Prevention*, *43*(3), 706-713.

Van Dam, A., 1997. Post-WIMP user interfaces. *Communications of the ACM*, *40*(2), pp.63-67.

Van Erp, J. B., & Van Veen, H. A. (2004). Vibrotactile in-vehicle navigation system. *Transportation Research Part F: Traffic Psychology and Behaviour*, *7*(4), 247-256.

Vanholme, B., Gruyer, D., Glaser, S., & Mammar, S. (2011, June). A legal safety concept for highly automated driving on highways. In *Intelligent Vehicles Symposium (IV), 2011 IEEE* (pp. 563-570). IEEE.

Várhelyi, A., Hjälmdahl, M., Risser, R., Draskóczy, M., Hydén, C., Taniguchi, S., & Ashouri, H. (2002, October). The effects of large scale use of active accelerator pedal in urban areas. In *International Cooperation of Theories and Concepts in Traffic safety (ICTCT) Workshop, Nagoya, Japan*.

Walker, G. H., Stanton, N. A., & Young, M. S. (2006). The ironies of vehicle feedback in car design. *Ergonomics*, *49*(2), 161-179

Walker, G.H., Stanton, N.A. and Young, M.S., 2001. Where is computing driving cars?. *International Journal of Human-Computer Interaction*, *13*(2), pp.203-229.

Weinberg, G. (2012) Tutorial: Voice and Multimodal Interaction in the Car. Tutorial. http://www.auto-ui.org/12/adjunct-proceedings/t1-voice-multimodal-nteraction-car-weinberg.pdf accessed on 11/11/16.

Weiner, J., 2007. Measurement: reliability and validity measures. Bloomberg School of Public Health, Johns Hopkins University, mimeo (Power Point Presentation) at http://ocw. jhsph. edu/courses/hsre/PDFs/HSRE_lect7_weiner. pdf. http://jae. oxfordjournals. org.

Wellings, T., Williams, M. A., & Pitts, M. (2008). Customer perception of switch-feel in luxury sports utility vehicles. *Food Quality and Preference*, *19*(8), 737-746.

Weng, F., Varges, S., Raghunathan, B., Ratiu, F., Pon-Barry, H., Lathrop, B., ... & Prieto, R. (2006, September). CHAT: a conversational helper for automotive tasks. In *INTERSPEECH*.

Whitaker, T. A., Simões-Franklin, C., & Newell, F. N. (2008). The Natural Truth: The Contribution of Vision and Touch in the Categorisation of "Naturalness". In *Haptics: Perception, Devices and Scenarios* (pp. 319-324). Springer Berlin Heidelberg.

Wickens, C.D., Lee, J., Liu, Y.D. and Gordon-Becker, S., 2013. *Introduction to Human Factors Engineering: Pearson New International Edition*. Pearson Higher Ed.

Wigdor, D., & Wixon, D. (2011). *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier.

Wilfinger, D., Meschtscherjakov, A., Murer, M., Osswald, S., & Tscheligi, M. (2011). Are we there yet? a probing study to inform design for the rear seat of family cars. *Human-Computer Interaction–INTERACT 2011*, 657-674.

Wilson, F. (1998) The Hand: How its use shapes the brain, language and human culture.

Windhager, S., Bookstein, F. L., Grammer, K., Oberzaucher, E., Said, H., Slice, D. E., & Schaefer, K. (2012). "Cars have their own faces": cross-cultural ratings of car shapes in biological (stereotypical) terms. *Evolution and Human Behavior*, *33*(2), 109-120.

Windhager, S., Slice, D. E., Schaefer, K., Oberzaucher, E., Thorstensen, T., & Grammer, K. (2008). Face to Face. *Human Nature*, *19*(4), 331-346.

Wiredu, E. 2015. Personal communication (email)

Wolcott, H.F., 1999. *Ethnography: A way of seeing*. Rowman Altamira.

Wynn, T., Richardson, J. H., & Stevens, A. (2013). Driving whilst using in-vehicle information systems (IVIS): benchmarking the impairment to alcohol.

Ylirisku, S., & Buur, J. (2007). *Designing with video*. Springer-Verlag London Limited.

You, H., Ryu, T., Oh, K., Yun, M. H., & Kim, K. J. (2006). Development of customer satisfaction models for automotive interior materials. *International Journal of Industrial Ergonomics*, *36*(4), 323-330.

Young, M. S., & Birrell, S. A. (2012). Ecological IVIS design: using EID to develop a novel in-vehicle information system. *Theoretical Issues in Ergonomics Science*, *13*(2), 225-239.

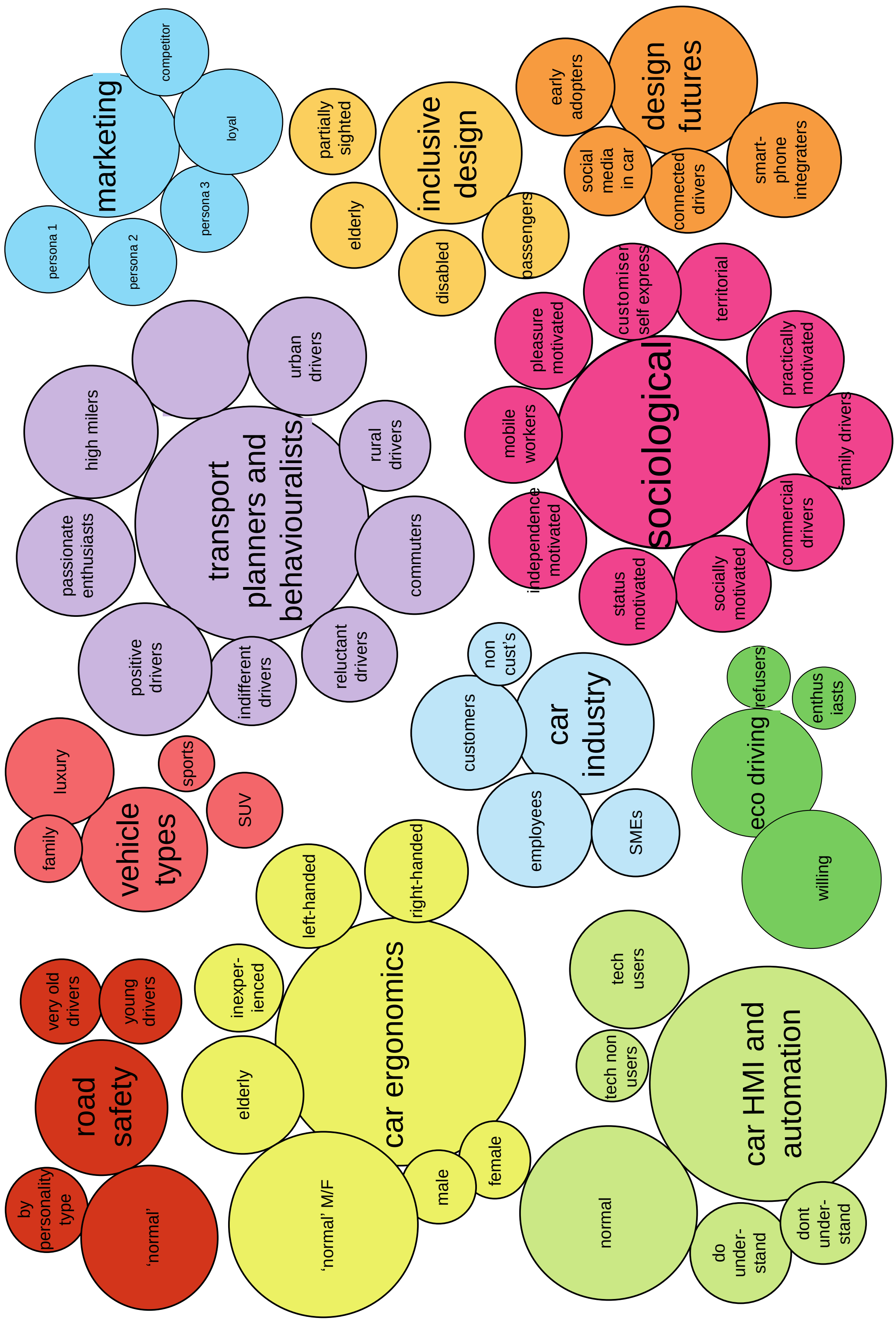
Young, M. S., Birrell, S. A., & Stanton, N. A. (2011). Safe driving in a green world: A review of driver performance benchmarks and technologies to support 'smart'driving. *Applied ergonomics*, *42*(4), 533-539.

Yun, M. H., You, H., Geum, W., & Kong, D. (2004, September). Affective evaluation of vehicle interior craftsmanship: systematic checklists for touch/feel quality of surface-covering material. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 48, No. 6, pp. 971-975). SAGE Publications.

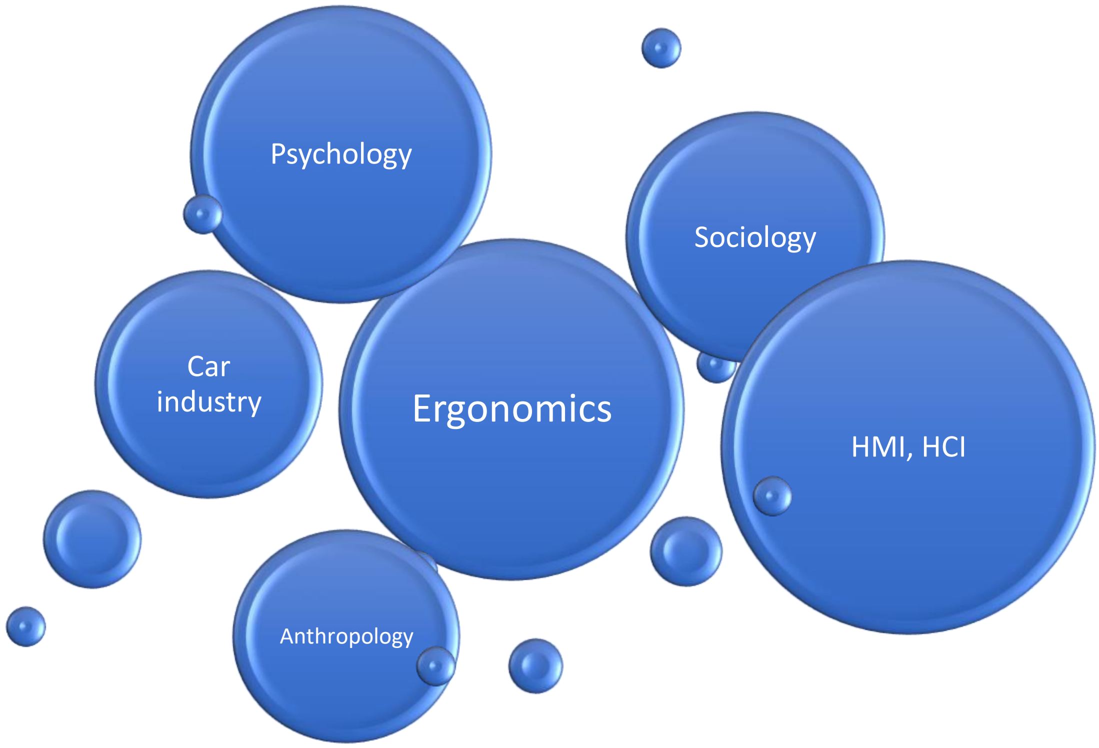
Zobl, M., Geiger, M., Bengler, K., & Lang, M. (2001). A usability study on hand gesture controlled operation of in-car devices. *Abridged Proceedings, HCI*, 5-10.

Appendix A

Infographics produced for literature review



A1: Stakeholder mapping before literature review



A2: Illustration of domains surveyed for literature review

Appendix B

Ethical approval and risk assessment (all studies)

Consent information (all studies)

Brunel University London
Uxbridge UB8 3PH
United Kingdom

www.brunel.ac.uk

30 March 2015

STATEMENT OF ETHICS APPROVAL

Proposer: Simon Ramm Student ID - 1139141

Dear Simon,

Project Title: Questionnaire based rating scale test for naturalness of secondary car controls

Under delegated authority from the College Research Ethics Committee, I have considered the application recently submitted by you. I am satisfied that there is no objection on ethical grounds to the proposed study.

Approval is given on the understanding that you will adhere to the terms agreed with participants and to inform me of any change of plans in relation to the information provided in the application form.

In addition, please provide notification to the College Research Office when the study is complete, if it fails to start or is abandoned.

Yours sincerely,

John Park
College Research Mananger
T +44(0)1895 266057| E john.park@brunel.ac.uk

Brunel University London
College of Engineering, Design and Physical Sciences

B1: College Approval Letter

Informed Consent Form [Study 1]

This informed consent form is for:

who we are inviting to participate in the research project titled 'Naturalness Framework for Person-Automobile Dialogue'.

This Informed Consent Form has two parts:

- **Information Sheet** (To share information about the research with you)
- **Certificate of Consent** (For signature, if you choose to participate)

Thank you for your participation

Simon Ramm
PhD Design Researcher at Brunel University

PERCEPTION ENHANCEMENT RESEARCH

PART OF THE HCDI Human Centred Design Institute

Brunel UNIVERSITY LONDON

Information Sheet [Study 1]

I. Introduction

You are being invited to take part in our research study that investigates the assessment and improvement of the 'naturalness' of interaction between drivers and their cars. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information. The research is being conducted as part of a PhD degree at Brunel University.

II. Purpose of the study

The aim of the study is to investigate factors that lead to perceived 'natural' interactions between a driver and his or her car with the aim of improving things in the future. As part of this research, in its early stages, we need to find out how ordinary drivers interact with their cars now, and what deeper meanings and expectations the car's components and controls might have for them.

III. Why you have been selected to participate

As part of the initial literature review for the project, certain key groups of car user (sometimes known as stakeholders) were identified as being academically and statistically relevant.

The criteria for participating in this study are (1) to be a car driver and aged 25-75 years (2) to be a car owner or to have regular use of at least one car (3) to be considered a representative of a particular group of stakeholder car drivers and (4) to drive more than 3,000 miles on an average year. No specialist knowledge whatsoever is required.

IV. Consent

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign the accompanying consent form.

V. Voluntary participation

Your participation in this research is entirely voluntary. You have the right to withdraw from the study at any time.

VI. Risks of participation

No significant risks or disadvantages to participants have been identified.

VII. Type of research intervention

This research will involve your participation in an audio-recorded interview (using a Dictaphone) that will last approximately 35 minutes. The interview itself will comprise of a series of short questions (28 in total) about cars, driving, and cars of the future. In order to maintain academic rigour and to facilitate analysis, it will only be possible to repeat the question rather than explaining it in other words, to avoid prejudicing your answers. Please just answer as best you can, using your interpretation of the words used in the question.

At the end of the interview do feel free to ask more information about the research. In the interests of academic rigour this information cannot be provided before or during the interview as it may prejudice your answers.

I plan to analyse the frequency of key words and themes given by the participants in relation to the interview questions asked, to identify patterns, correlations and insights. I may use a computer programme to assist in this.

VIII. What will happen to the results of the research?

The research will be published in my PhD thesis and be made available to our industrial sponsor. In addition they may also be published in academic papers resulting from the research, or other less formal writing such as blogs and magazine articles. The confidentiality arrangements are the same for all outputs and fully detailed below.

IX. Confidentiality

We will not be sharing information about you or your answers to anyone outside of the research team. The information that we collect from this research project will be kept confidential. After transcription of the audio recording you will only be identified by your short form biographical and car ownership data (for example "male, aged 60-65, SUV owner, automatic, 12,000 miles per year". Your name will never be used in any publication, nor even in discussion with other members of the research team, but direct quotes may be used in the writing up of the research. This is because qualitative research can be greatly enhanced by use of participants' own words. The writing up will enable the research to become a potential contribution to automotive knowledge.

The file name for your audio transcription will not contain your name, nor will the transcript itself.

While we are required to keep audio recordings for the duration of the PhD degree and for one year after, the recordings will be deleted after that time (2016-17). In any case, the digital recordings are not of a form which can be distributed or shared beyond the recording device itself.

X. Who to Contact

If you have any questions, please contact:

Simon Ramm
Mobile: +44 (0)7899 792319
Email: simonramm@yahoo.co.uk

This study has been approved by the School of Engineering & Design Research Ethics Committee of Brunel University.

Certificate of Consent [Study1]

I have received written explanation of the research and have also been given the opportunity to ask for clarification and/or further details should I wish.

I freely give my consent to take part in this research. I am a consenting adult over 18 years old and if I have any disability that will require adjustments to be made to the study, I will make the researcher aware of these prior to the arranged time and date of the agreed interview. I understand that I have the right to withdraw from the research at any time. My data will be stored securely and will also be made anonymous. The data may be used and published as part of academic research as a collective whole, but I still have the right to ask for my data to be removed should I so wish.

Signature of Participant: ....

Print Name of Participant: ....

Date: ....

If you have any questions, please contact the researcher, Simon Ramm.
Mobile: +44 (0)7899 792319 Email: simonramm@yahoo.co.uk

This study has been approved by the School of Engineering & Design Research Ethics Committee of Brunel University.

[Similar forms were produced for Studies 2, 3 and 4 – omitted from this copy]

Informed Consent Form [Study 2]

This informed consent form is for:

who we are inviting to participate in the research project with the working title 'Creative workshop about human centred design of future car controls'.

This Informed Consent Form has two parts:

- **Information Sheet** (To share information about the research with you)
- **Certificate of Consent** (For signature, if you choose to participate)

Thank you for your participation

Simon Ramm
PhD Design Researcher at Brunel University

Information Sheet [Study 2]

XI. Introduction

You are being invited to take part in our qualitative research study that investigates the feelings associated with interaction between drivers and their car, and operating its various controls, and how it might be in the future. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information. The research is being conducted as part of a PhD degree at Brunel University.

XII. Purpose of the study

The aim of the study is to investigate perceptions of interactions between a driver and his or her car with the aim of improving things in the future. As part of this research, in its early stages, we need to find out how ordinary drivers interact with typical car controls now, and what deeper meanings and expectations the car's controls might have for them.

XIII. Why you have been selected to participate

As part of the initial literature review for the project, certain key groups of car user (sometimes known as stakeholders) were identified as being academically and statistically relevant.

The criteria for participating in this study are (1) to be a car driver and aged 25 -75 years (2) to be a car owner or to have regular use of at least one car (3) to be able to express honest feelings and opinions to a high level, in the setting of a focus group. No specialist knowledge whatsoever is required. Your performance is not being monitored, only the issues you raise in the discussion.

XIV. Consent

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign the accompanying consent form.

XV. Voluntary participation

Your participation in this research is entirely voluntary. You have the right to withdraw from the study at any time. As a gesture of goodwill and in lieu of travel expenses, a voucher of your choice to the value of £20 will be sent after the research to all participants who remain for the whole duration.

XVI. Risks of participation

No risks or disadvantages to participants have been identified.

XVII. Type of research intervention

This research will involve your participation in an audio-recorded programme of practical creative exercises and discussions that will last approximately 2-3 hours. Most of it will be conducted around a table in a normal

Brunel laboratory setting, but some exercises may involve sitting in a safe parked car inside our laboratory, to get you closer to the perceptions and sensations of driving. You will also be asked to select various car switches or controls from a selection box in the workshop, and arrange them in various ways, and make notes or sketches if you wish. Please let me know if you have any impairment that would prevent you from fully participating and we will see what support can be arranged.

At the end of the interview do feel free to ask more information about the research. In the interests of academic rigour information cannot be provided before the interview as it may prejudice your answers.

Some still photography will be taken of anything you create or draw, and of you operating things with your hands; wherever possible your face will not be included in the frame.

I plan to analyse verbal themes given by the participants in relation to the questions and topic guides, to identify patterns, correlations and insights, and also visually analyse the paper outputs.

XVIII. What will happen to the results of the research?

The research will be published in my PhD thesis (without participants being named) and data may be made available to our industrial sponsor. In addition they may also be published in academic papers resulting from the research, or other less formal writing such as blogs and magazine articles. The confidentiality arrangements are the same for all outputs and more fully detailed below.

XIX. Confidentiality

We will not be sharing information about you or your answers to anyone outside of the research team and our industrial sponsor, and possibly a professional transcriber. The information that we collect from this research project will be kept confidential. After transcription of the audio recording you will only be identified by your short form biographical and car ownership data (for example "male, aged 60-65, SUV owner, 12,000 miles per year"). Your name will not be used in any publication, nor even in discussion with other members of the research team, but direct quotes may be used in the writing up of the research. This is because qualitative research can be greatly enhanced by use of participants' own words. The writing up will enable the research to become a potential contribution to automotive knowledge.

The file name for your audio transcription will not contain your name, nor will the transcript itself.

While we are required to keep audio recordings for the duration of the PhD degree and for one year after, the recordings will be deleted after that time (2016-17).

XX. Who to Contact

If you have any questions, please contact:

Simon Ramm
Mobile: +44 (0)7899 792319
Email: Simon.Ramm@brunel.ac.uk

This study has been approved by the School of Engineering & Design Research Ethics Committee of Brunel University.

B6: Information Sheet [Study 2]

Informed Consent Form [Study 3]

This informed consent pack has been prepared for you because we have invited you to participate in the research project titled 'Ethnographic observations of drivers driving ordinary journeys using secondary controls and systems'

This Informed Consent Form has two parts:

- **Information Sheet** (To share information about the research with you)
- **Certificate of Consent** (For signature, if you choose to participate)

Thank you for your participation

Simon Ramm
PhD Design Researcher at Brunel University

Information Sheet [Study 3]

XXI. Introduction

You are being invited to take part in our research study that investigates the assessment and improvement of interaction between drivers and their cars. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information. The research is being conducted as part of a PhD degree at Brunel University, London, England.

XXII. Purpose of the study

The aim of the study is to investigate 'natural' interactions between a driver and his or her car with the aim of improving things in the future. As part of this research, in its early stages, we need to find out how ordinary drivers interact with their cars now, and what deeper feelings, meanings and expectations the car's components and controls might have for them.

XXIII. Why you have been selected to participate

As part of the initial literature review for the project, certain key groups of car user (sometimes known as stakeholders) were identified as being academically and statistically relevant.

The criteria for participating in this study are (1) to be a car driver and aged 25-75 years (2) to be a car owner or to have regular use of at least one car (3) to be considered a representative of a particular group of stakeholder car drivers and (4) to drive more than 5,000 miles on an average year. No specialist knowledge whatsoever is required.

XXIV. Consent

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign the accompanying consent form.

XXV. Voluntary participation

Your participation in this research is entirely voluntary. You have the right to withdraw from the study at any time.

XXVI. Risks of participation

No significant risks or disadvantages to participants have been identified. The study will be purely observational and there will be no distracting flash photography or video. Only when the vehicle is safely stationery might you be asked questions. There is no payment incentive so having me present should not affect your insurance. Legally I will simply be a passenger in your car which will be covered by any normal insurance policy.

XXVII. Type of research intervention

This research will involve observing you driving a scheduled journey of around one hour. We would prefer that this be a journey you were planning to take anyway, rather than just for research purposes, so that the circumstances being observed are as 'natural' and realistic as possible.

The study will be purely observational and there will be no distracting flash photography or video. Only when the vehicle is safely stationery might you be asked questions about your car's controls. The researcher will take notes in a book. With permission, we may ask to take a still photograph when safe to do so, without a flash. I can sit in the passenger or rear seat, whichever you prefer.

At the end of the drive do feel free to ask more information about the research. In the interests of academic rigour this information cannot be provided before or during the interview as it may prejudice your answers.

XXVIII. What will happen to the results of the research?

The research will be published in my PhD thesis and be made available to our industrial sponsor. In addition they may also be published in academic papers resulting from the research, or other less formal writing such as blogs and magazine articles. The confidentiality arrangements are the same for all outputs and fully detailed below.

XXIX. Confidentiality

We will not be sharing information about you or your answers to anyone outside of the research team. The information that we collect from this research project will be kept confidential. After transcription of the audio recording you will only be identified by your short form biographical and car ownership data (for example "male, aged 60-65, SUV owner, automatic, 12,000 miles per year". Your name will never be used in any publication, nor even in discussion with other members of the research team, but direct quotes may be used in the writing up of the research. This is because qualitative research can be greatly enhanced by use of participants' own words. The writing up will enable the research to become a potential contribution to automotive knowledge.

Where photos are reproduced, your head and shoulders will not be included.

XXX. Who to Contact

If you have any questions, please contact:

Simon Ramm
Mobile: +44 (0)7899 792319
Email: simonramm@yahoo.co.uk

This study has been approved by the School of Engineering & Design Research Ethics Committee of Brunel University.

Informed Consent Form [Study 4]

This informed consent pack has been prepared for you because we have invited you to participate in the research project titled 'Driver-car interaction rating scale questionnaire'.

This Informed Consent Form has two parts:

- **Information Sheet** (To share information about the research with you)
- **Certificate of Consent** (For signature, if you choose to participate)

Thank you for your participation

Simon Ramm
PhD Design Researcher at Brunel University

Information Sheet [Study 4]

XXXI. Introduction

You are being invited to take part in our research study that investigates the interaction between drivers and their cars. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information. The research is being conducted as part of a PhD degree at Brunel University.

XXXII. Purpose of the study

The aim of the study is to investigate factors that lead to positive interactions between drivers and their cars with the aim of providing manufacturers with a checklist of design guidelines for improved automotive interaction design in the future. As part of this research, in its early stages, we need to find out how drivers interact with their cars at the moment and what feelings are evoked.

XXXIII. Why you have been selected to participate

As part of the initial literature review for the project, certain key groups of car user (sometimes known as stakeholders) were identified as being academically and statistically relevant. The criteria for participating in this study are (1) to be a car driver and aged 25-75 (2) to be a car owner or to have regular use of a car (3) to be a member of a stakeholder group.

XXXIV. Consent

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign the accompanying consent form.

XXXV. Type of Research Intervention

This research will involve your participation in a questionnaire administered in-person by the researcher, usually myself. The questionnaire will ideally be administered in your own car. Without turning on the engine, you will be asked to pick a suitable secondary (non driving) control or system in your car. The questionnaire will comprise 46 short multiple choice questions about this secondary control. It will take approximately 20 minutes. We plan to analyse the data obtained across all the participants using SPSS software to look for correlations, factors, patterns and averages.

XXXVI. Voluntary Participation

Your participation in this research is entirely voluntary. You have the right to withdraw from the study at any time.

XXXVII. Risks of participation

No risks or disadvantages to participants have been identified. We mitigate risks by insisting your car is parked safely in a car park, not on a steep incline, and the engine is not turned on. The handbrake and gear lever should not be adjusted nor should they be one of the systems you are answering questions about.

XXXVIII. What will happen to the results of the research?

The research will be published in the PhD thesis and be made available to our industrial sponsor. In addition they may also be published in academic papers resulting from the research, or less formal writing such as blogs and magazine articles. The confidentiality arrangements are the same for all outputs and detailed below.

XXXIX. Confidentiality

We will not be sharing information about you to anyone outside of the research team. The information that we collect from this research project will be kept confidential. Your name will not be used on the questionnaire, you will be identified only by your car type and system. Your name will never be used in any publication.

Following the gathering of data, an analysis of the results will be completed and written up which will then enable the research to become a potential contribution to automotive knowledge.

While we are required to keep paper data for the duration of the PhD degree and one year after, but it will be shredded thereafter.

XL. Who to Contact

If you have any questions, please contact: Simon Ramm
Mobile: +44 (0)7899 792319
Email: simonramm@yahoo.co.uk

This study has been approved by the School of Engineering & Design Research Ethics Committee of Brunel University.

University Research Ethics Committee RESEARCH ETHICS RISK ASSESSMENT AND MANAGEMENT			
Identified Risks	**Likelihood**	**Potential Impact/Outcome**	**Risk Management/Mitigating Factors**
Identify the risks/hazards present	High/Medium/Low	Who might be harmed and how?	Evaluate the risks and decide on the precautions, e.g., Health & Safety
Travel risks to location of research project:			
· Road/rail accident	Low		· Travel with companion
· Physical assault			· Awareness of options for mode of travel
			· Awareness of physical environment, e.g., alleyways, open spaces
Researcher:			· Researcher to be aware of health and safety policies of research location:
· Physical injury			o Fire bells
· Psychological harm			o Location of fire alarms & exits
Discussion of a sensitive topic in an interview has potential to cause distress to participant	Low	Participant:	· Offer to cease interview
e.g. of past car accident		· Psychological stress	· Signpost participant to external/internal support services
			exclude participants who have experienced serious car collisions via pre interview screening question
		Researcher:	
		· Anxiety about dealing with a complex situation	
Risk of explosion or accident in stationery car (during interviews)	Low	Participant and researcher	Only conduct on level ground and in gear/handbrake
			exit if odour of petrol
			conduct only off road on private land with minimal traffic movements (e.g. not car park)

Research Ethics Review Checklist

This checklist should be completed for every research project that involves human participation, the collection or study of their data, organs and/or tissue. It is used to identify whether a full application for ethics approval needs to be submitted. **Before completing this form, please refer to the University Code of Research Ethics and General Ethical Guidelines and Procedures.** The principal investigator or, where the principal investigator is a student, the supervisor, is responsible for exercising appropriate professional judgement in this review.The checklist must be completed before potential participants are approached to take part in any research.

Section I: Project details

1. Project title: 'Naturalness framework for person automobile dialogue'
2. Proposed start date: 15th June 2013 (ie this research start date)
3. Proposed end date: 15th August 2013

Section II: Applicant details

2. Name of researcher (applicant):	Simon Ramm

Section III: For students only

7. Supervisor's or module leader's name:	Professor Joseph Giacomin

Section IV: Description of project

Please provide a short description of your project:
The overall aim of this research is concerned with defining and assessing naturalness in driver interaction in the relevant literature, and in drivers' minds, to develop a framework for the different types of interaction between a driver and their car, what mental models and expectations exist, with the aim of producing guidelines and checklists for the automotive industry in order to have a greater chance of making driver interaction more natural in the future This may have other potential benefits such as improving user safety or satisfaction. The current research has recently identified a series of key user-stakeholder characteristics and subdomains that form part of the cultural domain of the automobile. As a pilot stage I have conducted 6 trial interviews to refine the semantics and question wording. This has shown promise that this ethnographic open interview style has potential to fill the gaps in the literature. The next stage of the research is to conduct a larger series of audio-recorded interviews, comprising of 25-30 driver stakeholders from various (automotive) subdomains to further identify drivers concepts and precepts of car interaction and what might feel more natural. While ethnographic in nature the questions are not personally challenging or uncomfortable as they focus entirely on driving and interaction with current and future cars. There is no deception or hidden agenda involved.

I propose to conduct all the interviews in person. I am prepared to travel to interviewees' place of home or work to conduct the interviews. Where possible I would like to conduct them in a stationary parked car (off-street for safety) which is one of the basic principles of ethnographic theory – ie to interview 'where the action takes place' and not in a laboratory environment. All interviews will be in the UK and I will travel by public transport or my own car, at my own risk, having had a drivers licence for 22 years and being fully insured.

All participants will be asked to sign an informed consent letter, attached. Please see the informed consent form for a more detailed summary of my proposed research and what I initially will send to the participants, for their consent and own information.

Following the gathering of data, an analysis of the results will be completed and written up (I plan to analyse the frequency of key words or concepts given by the participants in relation to the interview questions asked, to identify key themes and associations) which will then enable the research to become a potential contribution to knowledge.

Section V: Research checklist

Please answer each question by ticking the appropriate box:

	YES	NO
1. Does the project involve participants who are particularly vulnerable or unable to give informed consent (e.g., children, people with learning disabilities, your own students)? Etc deleted	☐	✔

Signed:
Date: 7th June 2013
Principal Investigator: Simon Ramm
Supervisor or module leader (where appropriate): Prof Joseph Giacomin

Appendix C

Early versions of the questionnaires used for pilot tests

46-question version of the questionnaire used for data collection in Chapter 7

List of all the 81 car systems tested in Chapter 7

C1: Longlist of 81 Questions used for pilot testing

Construct	Question/Item (to be asked after an actual interaction)	Neg anchor –Unnatural	Pos anchor + Natural
	The interaction felt	Unnatural	Natural
1	The design and layout of the control felt	Novel	Familiar
1	In response to my action, the car behaved	Unpredictably	Predictably
1	The car's response was	Surprising	Unsurprising
2	During the interaction it felt like	The car was fully in control	I was fully in control
2	In terms of the feedback the car gave me, I felt	Completely 'out of the loop'	Completely 'in the loop'
2	This car's 'locus of control' felt	Scattered	Focused
2b	Generally I would imagine operating this car would feel	Difficult	Easy
2b	Generally I would imagine operating this car would feel	Unsatisfying	Satisfying
2b	Generally I think this car would make me feel	Unsafe	Safe
2b	Generally I imagine this car would feel	Very separate from my life	Very integrated into my life
3	My communication with the road and surroundings felt	Very weak	Very strong
3	My communication with the car as a whole felt	Very weak	Very strong
3	My communication with the car as a whole felt	Artificial	Real
3	My communication with the car as a whole felt	Indirect	Direct
3	My relationship with the car felt	Distant	Close
4	Physically the control felt	Lightweight	Weighted
4	Physically the control felt	Loose	Tight
4	Physically the control felt	Imprecise	Precise
4	Physically the control felt	Clunky	Fluid
4	Physically the control felt	Flimsy	Robust
4	Physically the control felt	Hollow	Solid
4	Physically the control felt	Shiny	Matt
4	Physically the control felt	Hard to grip	Easy to grip
4	Physically the interaction felt	Delayed	Instant
4	After my input, the control	Stayed in the same place	Centred itself or returned to its original position
5	The interaction felt	Uncomfortable	Comfortable
5	The interaction felt	Public	Private
5	The interaction felt	Stressful	Relaxing
5	The control looked	ungainly	elegant
5/6	The cabin overall looks	Complicated	Simple

6	Compared to adjacent controls, the control I used was	Very close to other controls	Far apart from other controls
6	The control I used was	Small	Large
6	The control I used was	Designed for the convenience of engineers	Designed for the convenience of my body
6	The control I used was	Inconvenient	Convenient
6	The control I used was	Digital	analogue
6	The control I used was	Hard to find	Easy to find
6	Generally the controls in the car seemed	Cluttered	uncluttered
6	Compared to other controls in the car, the control I used was	Hard to distinguish	Easy to distinguish
6	Operating that control unintentionally (accidentally) would be	Very likely	Very unlikely
6	Compared to other controls in the car, the control I used was	Hard to discern by touch alone	Easy to discern by touch alone
6	The shape and movement of the control	Bore no resemblance to its function	Closely resembled its function
7	In terms of the visual attention it demanded, the interaction was	Very demanding	Very easy (*undemanding?)
7	I could have done the same interaction competently without using my eyes	Highly unlikely	Highly possible
7	The amount of visual feedback I had to take in was	Very high	Very low
7	The relative weighting of the types of feedback that indicated to me the car had understood my intended action were:	Heavily weighted towards the visual	Heavily weighted towards the audible, tactile and positional
8	Mentally, the action was	Very Demanding	Very Undemanding (*easy? Helpful?)
8	The interaction overall felt	Counter intuitive	Intuitive
8	The amount of decision-making required felt	Overwhelming	Minimal
8	If I was driving, this secondary interaction would have felt	Highly distracting	Not at all distracting * (again, not an opposite).
8	Alternative: If I had been driving, this secondary interaction would have	Made me fully engaged with this secondary task	Kept me fully engaged with the driving task
8	The interaction felt	Gimmicky	Essential
8	In terms of the geographical <u>mapping</u> of the control to its intended outcome (e.g. up, down, forward, back, left right etc.) the mapping felt	Arbitrarily or counterintuitively mapped	Logically mapped

8	Before operation, the control's action seemed to be	Very unclear	Very obvious
8	With experience, I could imagine myself unconsciously doing that interaction without having to think about it.	Disagree	Agree
9	The interaction made me feel	Subordinate to the car	Superior to the car
9	The car came across as	Machinelike	Humanlike
9	The car came across as a	Passive recipient	Active agent
9	the relationship between me and the car felt	One sided	Interactive
9	The relationship between me and the car felt	Informal	Formal
9	(*Imagining the car as a person) the car seemed	Unhelpful	Helpful
9	(*Imagining the car as a person) the car seemed	Nagging	Tolerant
9	(*Imagining the car as a person) the car seemed	Rude	Polite
9	(*Imagining the car as a person) the car seemed	Lacking in initiative	Proactive
9	(*Imagining the car as a person) the car seemed	Incompetent	Competent
	(*Imagining the car as a person) the car seemed	Uncommunicative	Very informative
10	The car seemed	incompetent	expert
10	The car acted like	It was selfish (*note "had no empathy" is unipolar not bipolar)	It had empathy with me
10	The car would probably	Act the same regardless of context	Adapt to context
10	The car gives the impression it	Does not know what is going on around it	Knows what is going on around it
10	The car would probably	Not remember any of my preferences	Remember all my preferences
10	In its dealings with me the car would be	Socially unaware	Socially aware
10	The car could probably project scenarios into the future and forsee eventualities	NOT AT ALL	COMPLETELY AGREE
11	If this car spoke to me the type of voice communication we used would be	Command style; learned vocab	Natural conversational style, no learned vocab.

11	If this car spoke to me it would recognize what I said to it	None of the time	All of the time
11	If this car spoke to me it would repeat messages	Rarely	Constantly
11	If this car spoke to me it would speak to me	Whenever it wanted	Only when spoken to
11	If this car spoke to me its messages would be	Fully articulated	Brief and to the point
11	If this car spoke to me its messages would	State the obvious	Assume I knew the obvious
11	If this car spoke to me its messages would be delivered	As soon as the car conceived it	Carefully timed to an opportune moment
11	If this car spoke to me its messages would	Be about any topic possibly relevant to me	Stick to its area of expertise (car related matters)
11	If this car spoke to me its tone of voice would be	Machinelike	Humanlike
11	If this car spoke to me its messages would be	Always delivered	Turned off if I wanted.

C2: Longlist of 46 Questions used for pilot testing

Name code:	Age:	**Car used for testing:**	**System used:**					**Usual car:**	
v 9	Please answer the following questions honestly but without over-analysing. You are asked to rate the control you've chosen on various 5-point scales below in terms of how much you think it agrees with the 'attribute' in Columns A or B. The two attributes are roughly 'opposites'. The mid-point represents 'neither one nor the other', the other points 'very' or 'somewhat'. You may also answer 'not applicable' in the right hand column if the question does not apply. Keep repeating the action to remind yourself how it feels. For example, if you felt the control's action was 'somewhat easy' you would mark 'X' as below:	Attribute 'A'	Very A	Somewhat A	Neither A nor B	Somewhat B	Very B	Attribute 'B'	OR question not applicable
E	*[Example question only] The control's action felt:*	*hard*				X		*easy*	
Now please start with the first set of questions, which is about how in control you feel									
1	The design and layout of the control is:	Novel						Familiar	
2	The car responds:	Unpredictably						Predictably	
3	When you do the action it feels like:	The car is fully in control						You are fully in control	
4	The action makes you feel you trust the car:	Not at all						Fully	
5	The system is controlled from:	Many locations						One location	
6	Operating the control feels:	Difficult						Easy	
7	The interaction feels:	Unnatural						Natural	
The next set of questions concerns what sense of connection you feel between you and the car or road.									
8	The communication between you and the car feels:	Indirect						Direct	
9	The communication between you and the car feels:	Artificial						Real	
10	The interaction feels:	Trivial						Serious	
11	The feedback of external conditions feels:	Weak						Strong	

The next questions concern the physical feel of the control in your hand.									
12	Physically, the control feels:	Lightweight						Weighty	
13	Physically, the control feels:	Loose						Tight	
14	Physically, the control feels:	Hollow						Solid	
15	Physically, the control feels:	Slippery						Tactile	
16	The control's response feels:	Delayed						Instant	
17	After your input, the control:	Stays in the same position						Returns to its original position	
The next questions are about the physical layout and visual demands of the car's controls									
18	The control is located:	Not at all where you expect						Where you expect	
19	The control is located:	Illogically						Logically	
20	The control's display (if any) looks:	Digital						Analogue	
21	Where the control is, it looks:	Cluttered						Uncluttered	
22	Operating it accidentally would be:	Highly likely						Highly unlikely	
23	The shape and movement the control:	Does not reflect its function at all						Closely reflects its function	
24	Visually, the interaction is:	Highly demanding						Not at all demanding	
25	If you needed to do it while driving you would:	Need to divert your eyes						Not need to divert your eyes	
The next questions are about how difficult or easy the action seems mentally.									
26	Mentally, the action is:	Highly demanding						Not at all demanding	
27	The interaction overall feels:	Counter intuitive						Intuitive	
28	The decision-making required is:	Overwhelming						Minimal	
29	If you were doing this task while driving, you would feel more engaged with:	This task						The (primary) driving task	
30	The input action required seems:	Completely unclear						Completely obvious	
31	With practice, you would probably operate it	Only with careful thought						Without thinking	

The next questions concern your imagined 'relationship' with the whole car, or its perceived personality.				
32	The interaction makes you feel like:	The car is in control of you		You are in control of the car
33	The car comes across as:	Passive		Active
34	The car comes across as:	Uncommunicative		Informative
35	(Imagining the car is a person) it seems:	Unhelpful		Helpful
36	(Imagining the car is a person) it seems:	Rude		Polite
The next questions are about what kind of 'intelligence' you think the car has, based on your interaction with it so far.				
37	The system seems:	Highly incompetent		Highly competent
38	The system seems to act:	For its own good		For your own good
39	The system seems:	Rigid		Adaptable
40	The system:	Forgets all your preferences		Remembers all your preferences
41	Overall the interaction felt	Unnatural		Natural
Only answer the next questions if the action uses your voice as an input or it spoke back to you.				
42	Its messages are (probably):	Never repeated		Frequently repeated
43	The system (probably) speaks:	Whenever it wants		Only when you ask it something
44	Its messages are (probably):	Articulated in full		Brief and to the point
45	Its messages are (probably):	About any topic that might be relevant to you		Only about car-related matters
46	Its tone of voice is:	Machine-like		Human-like

C3: List of all systems tested in Study 4 (Chapter 7)

System	Frequency (number tested)	Percentage
Windscreen wipers	9	11.1
Ventilation and heating system (not climate controlled)	9	11.1
Hifi/radio/CD/stereo	8	9.9
Climate control (automatic)	6	7.4
Electric windows	5	6.2
Trip computer/iDrive general	5	6.2
Indicators	5	6.2
Electric/automatic dim mirrors	3	3.7
Automatic headlights	3	3.7
Keyless start systems	3	3.7
Air vent swivel controls	2	2.5
Interior lights	2	2.5
Stop/start (fuel saving) systems	2	2.5
Mobile telephone integration/Bluetooth	2	2.5
Integrated GPS	2	2.5
Electric handbrake	2	2.5
Driving mode (e.g. Sport)	2	2.5
Electric seats	2	2.5
Electric boot/automatic tailgate	2	2.5
Keyless entry	1	1.2
Tyre pressure monitoring	1	1.2
Head up display control	1	1.2
Parking aids/camera/maps	1	1.2
Lane keeping assist	1	1.2
Electric sunroof	1	1.2
Heated seats	1	1.2
	81	100.0

Appendix D

Factor Analysis chosen solution summary

Partial view of correlation matrix (very large table not suitable for A4)

Factor Analysis scree plot.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumul %	Total	% of Variance	Cumul %	Total	% of Variance	Cumul %
1	7.923	46.607	46.607	7.923	46.607	46.607	4.659	27.405	27.405
2	1.622	9.543	56.150	1.622	9.543	56.150	3.491	20.535	47.940
3	1.148	6.752	62.901	1.148	6.752	62.901	2.544	14.962	62.901
4	.931	5.475	68.376						
5	.828	4.872	73.248						
6	.745	4.380	77.628						
7	.640	3.763	81.392						
8	.539	3.169	84.561						
9	.506	2.979	87.540						
10	.439	2.580	90.119						
11	.375	2.206	92.326						
12	.348	2.050	94.375						
13	.270	1.589	95.964						
14	.241	1.417	97.382						
15	.194	1.141	98.522						
16	.153	.898	99.420						
17	.099	.580	100.000						

Extraction Method: Principal Component Analysis.

D1: Factor Solution SPSS readout, chosen solution, total variance explained

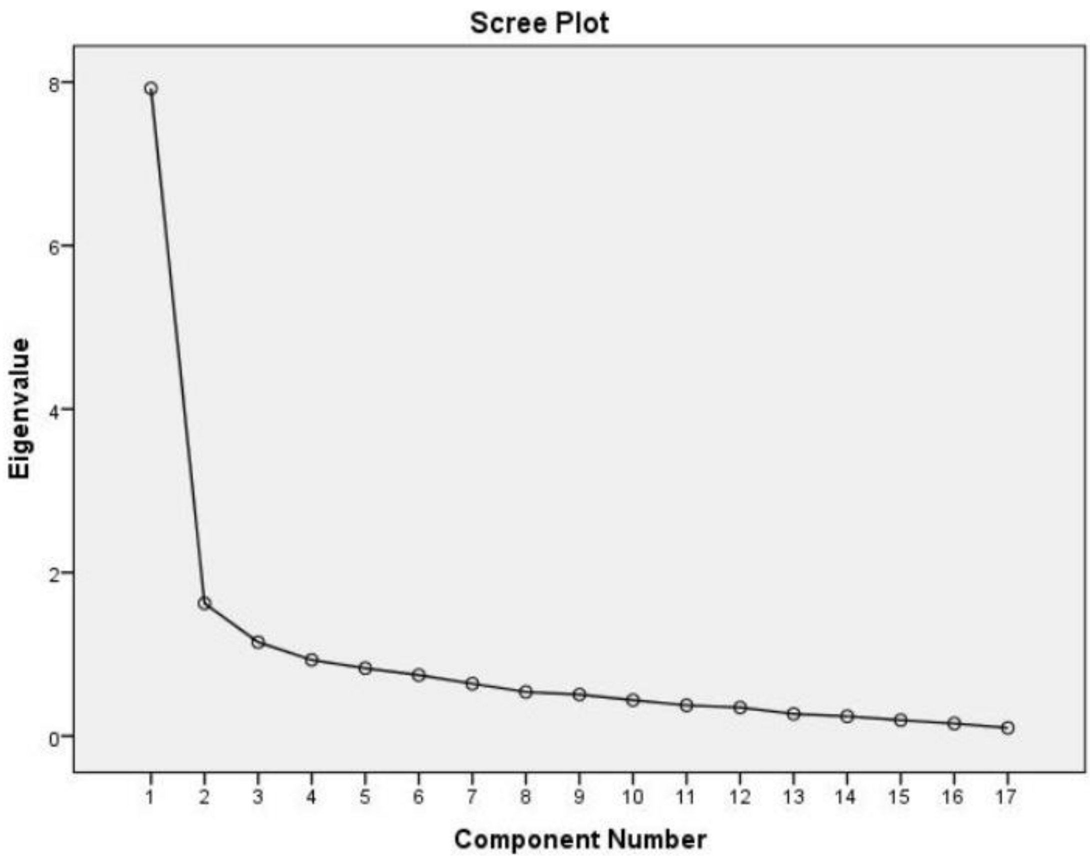
Factor Analysis

Correlation Matrix[a]

		Novel-Familiar	Notatall-Fully	Indirect-Direct	Artifical-Real	Trivial-Serious	Weak-Strong	Delayed-Instant	Illogic-Logic	Clut-Unclut	Reflects-notreflect	Divert-Notdivert	Unclear-obvious	Carcont-youcont	Uncom-Inform	Unhelp-Helpf	Rude-Polite	Incomp-Comp
Correlation	Novel-Familiar	1.000	.396	.435	.437	.183	.518	.392	.428	.261	.399	.366	.429	.424	.346	.374	.303	.295
	Notatall-Fully	.396	1.000	.522	.622	.383	.347	.341	.359	.332	.373	.391	.539	.747	.353	.497	.566	.616
	Indirect-Direct	.435	.522	1.000	.714	.345	.583	.607	.334	.259	.326	.372	.405	.544	.379	.503	.484	.492
	Artifical-Real	.437	.622	.714	1.000	.467	.551	.508	.381	.376	.477	.431	.607	.681	.428	.628	.584	.563
	Trivial-Serious	.183	.383	.345	.467	1.000	.315	.407	.237	.165	.321	.282	.239	.276	.256	.355	.330	.359
	Weak-Strong	.518	.347	.583	.551	.315	1.000	.556	.334	.264	.378	.376	.426	.451	.427	.497	.421	.360
	Delayed-Instant	.392	.341	.607	.508	.407	.556	1.000	.387	.124	.413	.522	.253	.303	.196	.269	.327	.361
	Illogic-Logic	.428	.359	.334	.381	.237	.334	.387	1.000	.332	.464	.386	.338	.360	.117	.276	.335	.171
	Clut-Unclut	.261	.332	.259	.376	.165	.264	.124	.332	1.000	.384	.232	.522	.417	.291	.444	.496	.418
	Reflects-notreflect	.399	.373	.326	.477	.321	.378	.413	.464	.384	1.000	.353	.512	.450	.314	.500	.410	.341
	Divert-Notdivert	.366	.391	.372	.431	.282	.376	.522	.386	.232	.353	1.000	.251	.494	.278	.342	.302	.423
	Unclear-obvious	.429	.539	.405	.607	.239	.426	.253	.338	.522	.512	.251	1.000	.607	.476	.684	.633	.461
	Carcont-youcont	.424	.747	.544	.681	.276	.451	.303	.360	.417	.450	.494	.607	1.000	.484	.620	.639	.591
	Uncom-Inform	.346	.353	.379	.428	.256	.427	.196	.117	.291	.314	.278	.476	.484	1.000	.789	.554	.566
	Unhelp-Helpf	.374	.497	.503	.628	.355	.497	.269	.276	.444	.500	.342	.684	.620	.789	1.000	.757	.711
	Rude-Polite	.303	.566	.484	.584	.330	.421	.327	.335	.496	.410	.302	.633	.639	.554	.757	1.000	.690
	Incomp-Comp	.295	.616	.492	.563	.359	.360	.361	.171	.418	.341	.423	.461	.591	.566	.711	.690	1.000
Sig. (1-tailed)	Novel-Familiar		.000	.000	.000	.051	.000	.000	.000	.009	.000	.000	.000	.000	.001	.000	.003	.004
	Notatall-Fully	.000		.000	.000	.000	.001	.001	.000	.001	.000	.000	.000	.000	.001	.000	.000	.000
	Indirect-Direct	.000	.000		.000	.001	.000	.000	.001	.010	.002	.000	.000	.000	.000	.000	.000	.000
	Artifical-Real	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	Trivial-Serious	.051	.000	.001	.000		.002	.000	.017	.070	.002	.005	.016	.006	.011	.001	.001	.001
	Weak-Strong	.000	.001	.000	.000	.002		.000	.001	.009	.000	.000	.000	.000	.000	.000	.000	.000
	Delayed-Instant	.000	.001	.000	.000	.000	.000		.000	.135	.000	.000	.011	.003	.039	.008	.001	.000
	Illogic-Logic	.000	.000	.001	.000	.017	.001	.000		.001	.000	.000	.001	.000	.149	.006	.001	.064
	Clut-Unclut	.009	.001	.010	.000	.070	.009	.135	.001		.000	.019	.000	.000	.004	.000	.000	.000
	Reflects-notreflect	.000	.000	.002	.000	.002	.000	.000	.000	.000		.001	.000	.000	.002	.000	.000	.001
	Divert-Notdivert	.000	.000	.000	.000	.005	.000	.000	.000	.019	.001		.012	.000	.006	.001	.003	.000
	Unclear-obvious	.000	.000	.000	.000	.016	.000	.011	.001	.000	.000	.012		.000	.000	.000	.000	.000
	Carcont-youcont	.000	.000	.000	.000	.006	.000	.003	.000	.000	.000	.000	.000		.000	.000	.000	.000
	Uncom-Inform	.001	.001	.000	.000	.011	.000	.039	.149	.004	.002	.006	.000	.000		.000	.000	.000
	Unhelp-Helpf	.000	.000	.000	.000	.001	.000	.008	.006	.000	.000	.001	.000	.000	.000		.000	.000
	Rude-Polite	.003	.000	.000	.000	.001	.000	.001	.001	.000	.000	.003	.000	.000	.000	.000		.000
	Incomp-Comp	.004	.000	.000	.000	.001	.000	.000	.064	.000	.001	.000	.000	.000	.000	.000	.000	

a. Determinant = 1.608E-5

D2: Correlation Matrix (partial view only because of size)



D3: Scree plot used in determining factor solution in principal components analysis