RESOLVING THE MISALIGNMENT BETWEEN CONSUMER PRIVACY CONCERNS AND UBIQUITOUS IS DESIGN: THE CASE OF USAGE-BASED INSURANCE

Completed Research Paper

Johannes Paefgen
University of St. Gallen
Chair of Operations Management
Dufourstrasse 40A
9000 St. Gallen, Switzerland
johannes.paefgen@unisg.ch

Thorsten Staake
ETH Zurich
Chair of Information Management
Weinbergstrasse 58
8092 Zurich, Switzerland
tstaake@ethz.ch

Frédéric Thiesse
University of Wuerzburg
Chair of Information Systems Engineering
Josef-Stangl-Platz 2
97070 Wuerzburg, Germany
frederic.thiesse@uni-wuerzburg.de

Abstract

Ubiquitous IS enables novel services and business models, yet require a careful balancing of consumer privacy concerns (PC) – induced by the provision of particular sensors and information types – with functional performance in order to maximize acceptance. For the exemplary case of Usage-based Insurance (UBI), this paper presents a design science approach to the mitigation of PC under parallel consideration of functional system performance. Based on long-term location trajectories from 1'600 vehicles, we assess the predictive power of emulated system designs that substitute location information, presumably the most privacy sensitive type of information in current UBI designs. We find that there are substantial grounds to challenge prevalent design paradigms in UBI and infer general insights from this example for IS researchers and IT professionals, who, when seeking to improve system privacy, often focus on privacy-enhancing technologies instead of considering the socio-technical context of ubiquitous IS.

Keywords: Ubiquitous computing, Privacy, Logistic regression, Information technology, Design science

Introduction

The advent of wireless sensors, location systems, and other ubiquitous computing technologies facilitates the cost-effective capture and processing of fine-grained data from the physical world (Fleisch and Thiesse 2007). The availability of such data and related context information on real-world events enables a broad range of novel services, yet at the same time raises privacy concerns (PC) among citizens due to the increased visibility of their personal habits and behaviors (Fano and Gershman 2002; Thiesse 2007; Iachello and Hong 2007; Hong et al. 2008). IS research has shown that PC are an important determinant of consumers' technology acceptance in applications that involve the use of privacy-sensitive data (Angst and Agarwal 2009; Dinev and Hart 2006; Sheng et al. 2008). In order to facilitate a widespread adoption of respective applications, it is crucial to balance the benefits of functional features of ubiquitous IS – which may increase perceived usefulness or ease of use – against their cost in terms of the PC they raise (Park 2009; Zuo 2010).

A prominent case of this critical trade-off is the vividly discussed concept of usage-based motor insurance (UBI). Under UBI contracts, insurance providers recalculate premiums in regular intervals – every month, for instance – based on individual driving patterns. This approach brings forward substantial benefits such as reduced information asymmetry between insurers and policyholders, has been shown to motivate safer driving (Bolderdijk et al. 2011), and also renders the substitution of discriminatory rate factors such as gender and nationality feasible. In spite of these promises, the market penetration of UBI has remained behind expectations with less than 1% of policies in Europe and North America. Among others, the slow diffusion rate of UBI has been attributed to PC among potential customers (Filipova-Neumann and Welzel 2010), which pose a major challenge to insurance IT professionals considering the introduction of UBI systems. Prior research has addressed this issue by proposing a variety of privacy-enhancing designs for their implementation (Coroama 2006; Duri et al. 2002; Iqbal and Lim 2010; Popa et al. 2009; Troncoso et al. 2007). However, these contributions do not question the design paradigm of location information as the predominant pricing criterion – a surprising finding considering the evidence in the IS privacy literature that indicates PC to be conditional on information type and particularly elevated for location information (Phelps et al. 2000).

As Langheinrich (2001) has postulated in his principle of *collection limitation*, ubiquitous IS design should incorporate only a minimum of privacy-sensitive data that suffices to a given system objective. In this view, a minor reduction in system performance is acceptable if it yields significantly improved acceptance through the abandonment of particularly privacy-sensitive information. In this paper, we argue that location information is substitutable by alternate information types which not only exhibit higher acceptance among consumers, but also yield comparable results in terms of predictive power for insurance ratemaking. While the mitigation of PC often increases the technological complexity of ubiquitous IS, for which UBI is a prototypical example, our approach shows how collection limitation can help to reduce complexity and create 'privacy-by-architecture' (Spiekermann and Cranor 2009). Following the call by Belanger and Crossler (2011) for more IS privacy research that pursues a "design and action research perspective with an eye towards actual implementation", we build upon established frameworks of design science research (Hevner et al. 2004; Peffers et al. 2007) in the proposition of a novel UBI system design that obviates location information. In order to evaluate the consistency of this IS artifact with business requirements – that is, the prediction performance toward insured risk – we emulate our design and provide a comparative assessment against conventional, location-based UBI in a logistic regression model. Our evaluation builds on an extensive dataset of GPS trajectories collected from 1'600 vehicles, of which 600 were involved in an accident during the observation period. We expect our research to initiate further studies on the substitution of location information in other ubiquitous IS applications and prompt managers as well as IT professionals to challenge established design paradigms.

The remainder of this paper is organized as follows. In the next section, we give a brief introduction into UBI system design and discuss information-type specific PC among potential customers. Based on this grounding, we propose a conceptual artifact design together with an appropriate evaluation procedure and discuss our empirical results. The paper closes with implications for further IS research and a set of guidelines for UBI stakeholders including insurers, hardware suppliers, and service providers.

Background and Motivation

Usage-based Insurance

Dynamic premiums and usage-based payment models for insurance have been addressed by IS research areas as diverse as ubiquitous systems (Fleisch and Thiesse 2007), business value of IT (Kohli and Grover 2008) and demand heterogeneity (Sen et al. 2009). Essentially, UBI systems enable insurance tariffs that (1) more closely reflect actual risk exposure under varying external conditions and (2) are adaptive over time, thereby yielding risk-minimizing incentives for policyholders. In consequence, information asymmetry between insurers and policyholders is reduced, which mitigates adverse selection and moral hazard (Chiappori et al. 2006). In the domain of UBI for automobile insurance – often also referred to as pay-as-you-drive (PAYD) insurance – existing studies have reported a significant impact on safe driving behavior among young adults, for instance (Bolderdijk et al. 2011). In addition, UBI has been associated with macroeconomic benefits such as insurance affordability, consumer surplus, improved traffic safety, and reduced externalities in the literature. By substituting conventional lump-sum premium payments with flexible rates, mobility costs are more adequately allocated (Litman 1997). Not least, insurance pricing based on objectively measured risks may replace potentially discriminatory rate factors such as gender or nationality (Buzzacchi and Valletti 2005), a recent insurance regulation issue in many countries.

A conceptual model of an information system for UBI is given in Figure 1. On the physical level, the model distinguishes between two separate functional units, namely a sensor unit installed on the client side in an insured vehicle and the insurance provider's backend system, which may be partially outsourced to a third-party system provider. The sensor unit in the vehicle records risk-related variables and transmits them periodically to an insurance provider's data repository, typically over a wireless communication network. Based on long-term records of these data as well as historical data on insurance claims associated with specific vehicles, insurers implement an actuarial model that estimates accident risk and expected claims cost in the form of a pure premium (Frees 2008). Using these estimates, a premium calculation module determines the actual premium to be billed to the policy holder, taking into account further pricing factors such as discounts, rewards, and competitive tariffs in the market.

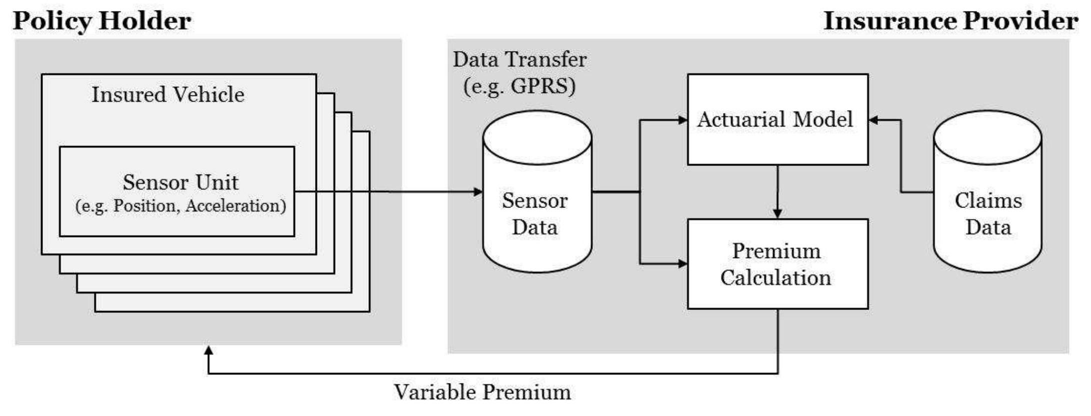

Figure 1. Conceptual model of a ubiquitous IS for UBI contracts

The overview of European UBI offerings given in Appendix A revealed that nearly all UBI system designs are at least to some extend location-based and require the installation of sensor units with GPS functionality. GPS trajectories are a rich source of accident risk-related information, although published results on this relationship remain sparse due to the reluctance of insurance providers to grant access to collected data. Commonly seen as the most indicative variable for accident risk used in UBI is vehicle mileage, which is also an established rate factor in conventional motor insurance. Further GPS-derived variables that have an empirically confirmed influence on accident risk are time of day, vehicle location, and average velocity (Jun et al. 2010; Paefgen et al. 2011; Progressive Insurance 2005). Accelerometers are a second, less relevant type of sensors in UBI systems and provide longitudinal and lateral accelerations that allow for an assessment of individual driving style (Toledo et al. 2008). To our knowledge, no UBI system implemented today makes use of proprietary vehicle sensors. Odometers, for instance, may also pose a reasonable source of vehicle mileage data, but accessing odometer data via external devices is a complex task that requires systems to be adaptive to various vehicle models which limits a widespread adoption.

Information Type-dependent Privacy Concerns

The privacy-sensitive nature of location information has motivated a large number of studies in IS privacy research. The limited exploitation and diffusion of location-based services has been attributed to a significant extent to PC, as location information allows for the association of lifestyle habits, behaviors, and movements with the consumer's personal identity (Xu and Teo 2004, 2005). Researchers have also investigated location privacy as an antecedent of technology acceptance (Xu et al. 2005; Yun et al. 2011; Zhou 2011). However, although it has been suggested that PC are information type dependent (Phelps et al. 2000), there is a lack of IS research that examines PC associated with location information in comparison to other types of information, or explores hierarchical structures of PC that would allow implementation-oriented researchers and IT professionals to prioritize privacy-enhancing measures (McParland and Connolloy 2008). According to the concept of privacy calculus, individuals evaluate privacy risks against benefits in order to decide whether to disclose personal information to complete a transaction (Dinev and Hart 2006). Though privacy calculus has found broad adoption in the context of location-based services (Pee 2011; Xu et al. 2009), it remains an open question how differential assessments are made in the presence of technological alternatives.

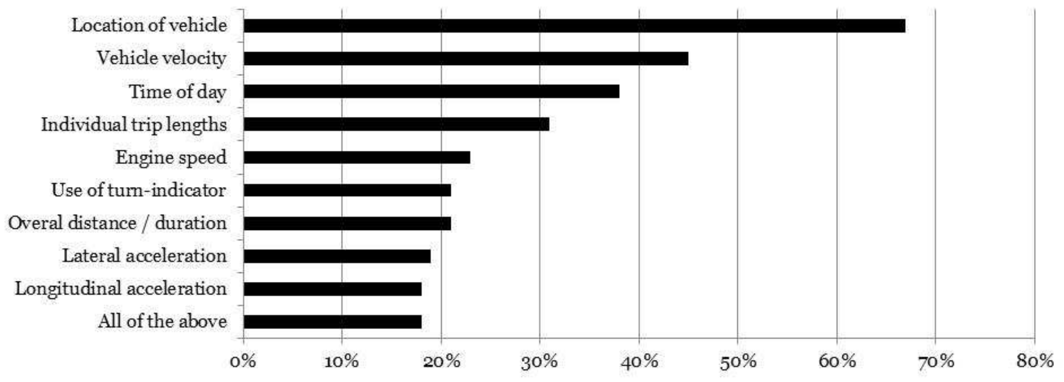

Figure 2. Information types policy holders are *not* willing to share with an insurance company, percentage of survey respondents (Mayer, 2012)

A recent technology acceptance study conducted by our research group confirmed the significant influence ($p<0.01$) of perceived privacy on intention to adopt for the specific case of UBI and yielded valuable insights into its information type dependency (Mayer, 2012). Respondents were requested to answer which types of vehicle sensor data they were *not* willing to share with an insurance provider. As depicted in Figure 2, more than two thirds (67%, rounded) of respondents considered vehicle location the most sensitive type of information, followed by vehicle velocity (45%), time of day (38%) and individual trip lengths (31%). Overall vehicle usage – presumably the most relevant UBI variable – finds comparatively broad acceptance with only 18% of respondents objecting to its use. Although these figures may vary according to regional and cultural preferences, we claim that our findings are generalizable on an ordinal scale and call for other researchers with different cultural background to put this hypothesis on a test bench. For the remainder of this paper, it is taken as premise that the privacy-sensitive nature of location information is at least a significant factor in the broad rejection of UBI models on the part of consumers.

Privacy-enhancing Design of UBI Systems

A plethora of research has addressed the privacy-enhancing design of UBI systems and the issue of location information in particular. Duri et al. (2002) presented a data protection framework in which the two system entities, vehicle and insurance backend, are separated by a "blackboard" and communicate only via data protection management agents. In this way, the premium calculation module only has access to aggregated data constructs instead of positioning data itself. This approach is extended by Iqbal and Lim (2010), who propose a system design in which policy holder and insurer communicate only in the form of a digital currency and the premium calculation module is implemented in the vehicle system. In this scenario, the actuarial modeling module from Figure 1 has only access to aggregated and anonymized data from a larger group of insured vehicles. Both propositions have evident weaknesses with respect to sensor manipulation or man-in-the-middle attacks. Moreover, they are difficult to audit – an important requirement for a system that determines insurance premium payments – and placing confidential algorithms for tariff calculation on distributed systems may be seen as a threat from an IT security perspective. Coroama (2006) has therefore suggested a more elaborated system that incorporates a third party service provider and sophisticated encrypted communication protocols. Troncoso et al. (2007) have also supported the notion of data aggregation and evaluation on the client side, and contributed a detailed analysis of legal implications, implementation costs and auditing aspects under this scenario. Among others, their paper provides a review of data gathering and transmission methods among 16 existing UBI offerings. A very extensive treatment of the matter from a technological perspective was recently carried out by Popa et al. (2009), who separated the UBI premium calculation process into vehicle registration, driving and reconciliation phases with distinct tasks carried out in each phase. The authors supplemented a mathematical analysis of encryption methods and protocols.

Though these studies propose important design improvements to enhance UBI privacy, they are of a technology-oriented nature and difficult to explain to potential customers, thus unlikely to increase acceptance. Insurance customers cannot be expected to develop a full understanding of the data structure, encryption and access restriction mechanisms and will remain skeptical toward the inclusion of location information. Furthermore, researchers have pointed out that the involvement of third parties as privacy guardians is of limited effectiveness in mitigating PC and privacy risk perception (Xu et al. 2008), which is intuitive due to the increased number of possible leaks or access violations. Another critique of the surveyed literature regards the lack of consideration bestowed to the business requirements in system design (Zuo 2010). For actuarial ratemaking, UBI system design should provide data of high predictive performance regarding the probability of accident events for an insured vehicle. Moreover, this predictive performance should be optimized subject to cost constraints, and especially the client side components of a UBI system should be scalable and yield low operational expenses. Though the technical design variations reviewed above claim to be of equivalent functional value, most of them are likely to incur additional system costs. Also, we found no empirical evidence on determinants of UBI system performance that would allow for inferring comparative evaluations. These unresolved questions leave insurance IT professionals without consistent advice and, due to a lack of rigorously tested reference designs, contribute to the fact that many insurers reject UBI as a whole.

Design Science Approach to UBI Privacy Enhancement

As Prince and Barrett (2005) have stipulated, privacy concerns should be incorporated early on in technology innovation and design processes, due to the difficulty of altering already developed and partially adopted technologies. Available UBI systems in the market today, however, appear to have been optimized from a performance rather than from an acceptance perspective. Furthermore, proposed design alterations have focused on technological aspects and have not addressed the tradeoff between PC and possible decreases in functional performance. We thus conclude that there is a research gap regarding a holistic design approach that considers UBI systems in their larger socio-technical frame and brings together information-type dependent PC and actuarial performance. A suggested IS research perspective on ubiquitous UBI system design is depicted in Figure 3.

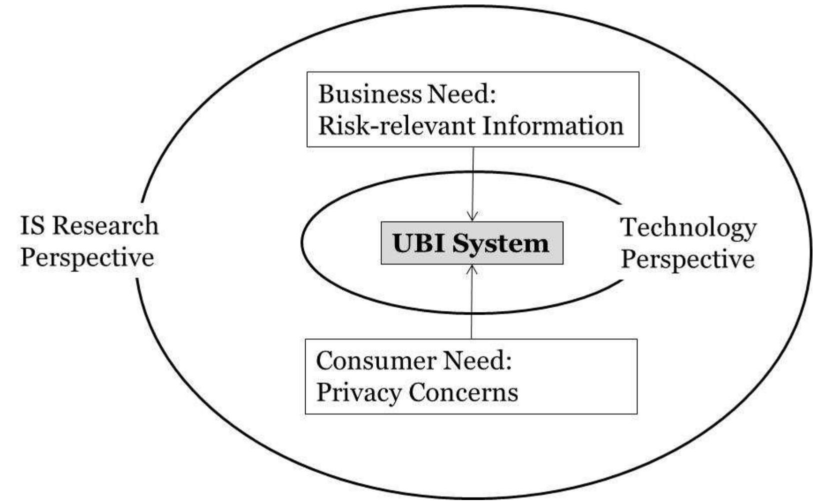

Figure 3. An IS Research perspective on UBI should incorporate PCs and the prediction of risk as design objectives

A recent comprehensive reflection on privacy research in IS by Belanger and Crossler (2011) suggested a new emphasis on design and action research methodologies in future studies on the subject. According to Hevner et al. (2004), the design science paradigm seeks to solve problems by extending existing boundaries of existing capabilities through the creation of novel artifacts. Artifacts are rarely full-grown information systems, but can be conceived as "innovations that define ideas, practices, technical capabilities, and products through which the analysis, design, implementation, and use of information systems can be effectively accomplished". We orient our research on the DSRM process model by Peffers et al. (2007). As elaborated in the previous sections, there is substantial grounding in both the application domain and in theoretical contributions to pursue a design science approach in the subject of UBI system privacy. While we consider the isolated technology perspective on UBI systems sufficiently covered, we challenge the dominant design paradigm of location information in the light of reported PC by potential customers. Consequently, we define the objectives of our proposed artifact in accordance with the second DSRM tier as follows. Our proposed novel UBI system design should abstain from the use of location information (1), be of lesser or equal implementation costs (2) and, most importantly, match or outperform existing solutions in terms of actuarial performance, i.e., the prediction of accident probabilities (3). Particularly the third objective is non-trivial and demands a careful and rigorous evaluation of our design. As we are concerned with the collection procedures of UBI data, our artifact contribution is of type *method*.

Location information is required in the first place in order to correctly determine the *driven distance* of a vehicle. To take into account small changes in driving direction and arrive at a reasonable precise measure of driven distance, high resolution position increments are aggregated. Distance, or mileage, is in fact one of the most established rate factors in conventional insurance models. In UBI systems, it is available at a higher level of precision, together with corollary information regarding situational factors, most prominently time of day, that are also indicative of accident risk and may serve to fine-tune the weighting with which a driven distance is included in premium calculation. For determining driven distance, location information appears without alternative due to the difficulties of interfacing proprietary vehicle electronics. If one inquires, however, which variable might be, at least to some extent, correlated to driven distance and allowing for alternate means of acquisition while maintaining its predictive power, driving *duration* is a reasonable candidate. While driving duration is also accessible from positioning data, one can conceive different solutions for its recording. For this purpose, we propose to employ a device that monitors a vehicles power network to determine its current ignition state. Alternatively, an accelerometer with an appropriately set threshold can be used that is triggered by vehicle movement activity. The two design concepts are visualized in Figure 4.

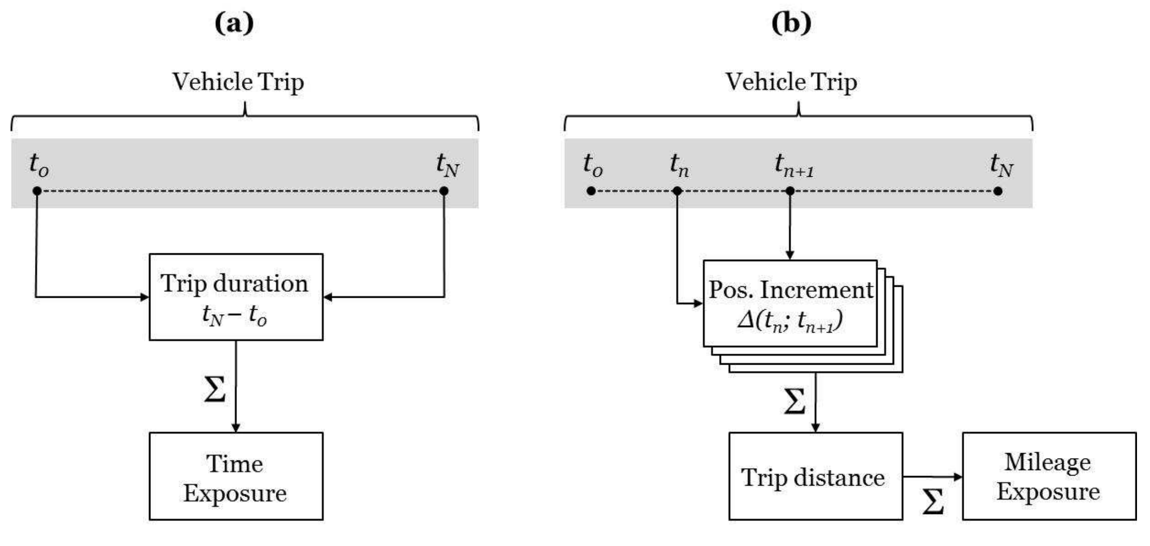

Figure 4. Ignition trigger (a) versus position increment (b) exposure aggregation

In spite of the minor technical modification a shift in UBI system design from location information-based distance to driving duration measurement seems to be, it in fact questions current procedures and methods and poses a significant challenge of proving the equivalence of driving distance and duration. Furthermore, although a trigger based system enables the determination of time of day conditions, it is not capable of recording vehicle velocity or location-related factors such as road type. Thus, the main contribution of our paper regards the evaluation of the proposed design as the next methodological step in the DSRM process. In accordance with Peffers et al. (2007), we proceed to establish relevant metrics and analysis techniques in the following section.

Empirical Evaluation

Methodology

In order to assess the suitability of the discussed exposure variables for variable premium calculation in a UBI context, we proceed with a comparison of their predictive power with respect to accident involvement probability. As the behavior of an ignition-trigger exposure sensor in a vehicle can be emulated from location information, we construct both distance and duration exposure together with time of day influence factors and compare the resulting variable sets. For this purpose, a model is required that employs monthly exposure as an input variable and assigns an accident likelihood to a vehicle ex-post, which in turn forms the scaling factor for a base premium.

We emphasize that this is not a standard binary classification task, as classes are not clearly assignable to objects and vary over time: Vehicles that exhibit very low exposures may still be involved in traffic accidents, and vehicle mileage and driving duration can be altered purposefully or due to unrelated influences. In fact, we expect a significant fraction of the differences between accident-free vs. involved drivers to be attributed to factors not accessible from location information, most prominently the conventional driver-specific rating factors that may be expressed by a constant component in the insurance tariff. We suggest multivariate logistic regression as an appropriate model. Logistic regression estimates an odds ratio that has some intuitive similarity to accident involvement probability. It suits the structure of our problem with continuous explanatory and binary dependent variables, and features established goodness-of-fit measures by which variable sets can be compared.

Sample Description

We obtained data from the database of a major European UBI provider that currently comprises more than 1.0m vehicles. Each vehicle is equipped with an on-board unit that includes a GPS sensor and wireless transmission capabilities. During vehicle operation, position updates were carried out every couple of seconds and aggregated on the device level to reduce costs of transmission and storage. For aggregation, the system calculated travelled distance from incremental position updates and generated new data entries every 2’000 meters. Next to a vehicle’s latitude and longitude, data points consisted of a time stamp, ignition status of the vehicle, and driven distance since the previously generated data point. This distance could in some cases exceed the 2’000 meter interval if no position update was available for some time, for example, owing to signal obstruction. Through straightforward computations, we extended raw data points to include the elapsed time since the last update, which in turn allowed us to compute the average velocity for the previously driven distance. In addition, the system inferred a road type indicator from data point locations, which distinguished urban roads, extra-urban roads, and highways. Start and end locations of vehicle trips were available from data points generated upon changes of the vehicle ignition status (i.e. engine start and switch off).

In its entirety, the database is computationally intractable by means available to us, so that we resorted to a randomized sampling procedure. We obtained reference samples for “high” and “low” risk driving patterns as follows. We randomly drew a sample of 600 vehicles that had an accident in the year of 2008, which contained six months of location data prior to the accident event. We used stratified sampling to achieve an even distribution of accident events over the year, so that one-twelfth, i.e. 50 vehicles, shared the same month in which the accident occurred. By sampling an equal amount of accident events for each month, we hope to eliminate the effect of seasonal variations on accident frequencies in our analysis. No location data beyond an accident event were included in the sample, as previous work has reported strong variations in driving patterns in the aftermath of an accident (Mayou et al. 1993). As a baseline, we furthermore randomly drew a second sample of 1000 vehicles from the data pool with twenty-four months available location data without accident-involvement throughout this period, reaching from July 2007 to June 2009. For privacy reasons, no driver particulars of any kind were included in the sample.

A number of vehicles were eliminated from both samples due to the following reasons:

- Further accident events in the six-month observation periods of accident-involved vehicles that would affect vehicle usage,
- Errors in data recording or storage that rendered the resulting log files infeasible to process.
- Failure of GPS sensors over prolonged periods, so that no location data was available for certain vehicles while ignition status indicated vehicle use, and
- Non-continuous GPS readings that resulted in excessively long travelled distances.

These instances were identifiable without doubt and resulted in a reduction of the accident-involved sample by 17 vehicles to 583 and of the accident-free sample by 16 vehicles to 984. No further elimination of outliers was undertaken, since we argue that due to the large sample size their effect on the inference statistics reported below is negligible. Both samples combined cover approximately 45.7×10^6 kilometers driven distance in 1.0×10^6 hours of vehicle operation.

Explanatory Variables

Figure 5 depicts the distribution of the two exposure variables in our sample, separated by the two vehicle groups. Differences in distribution means between accident-free and accident-involved drivers are highly significant for both driving duration ($t(1'565) = 15.94$, $p < .001$) and mileage ($t(1'565) = 15.192$, $p < .001$) and amount to 32.13 hours and 1'596.98 km, respectively. The distributions for accident-free drivers have a higher skewness and a higher kurtosis than for the accident involved drivers. We interpret this observation as a tendency of drivers in the accident-free vehicle group to exhibit more constant and regular driving patterns, such as a higher frequency of intra-urban trips of lesser distance. Irregularities in the long tails of both distributions may be attributed to the varying case number, while the overall trend is clear.

As a preliminary indicator for the similarity of both time-based and distance-based exposure metrics we report the results of a regression analysis. Figure 6 depicts the monthly average driving duration on the x-axis versus the monthly averaged driven distance on the y-axis across the sample in log-log scale – driving exposure typically follows a log-normal distribution (Rai and Singh, 2006). A linear regression with zero-fixed offset ($R^2 = .922$) yields a slope of 43.9 km/h, corresponding to the vehicle velocity that best explains the observed relationship between time and distance. When differentiated according to accident and accident-free vehicle groups, the linear coefficients are approximately equal, so that we claim similar predictability of distance from time and vice versa in both groups. On a side note, we observe that for our dataset, vehicle velocity does not seem to be globally indicative of accident risk as regression lines would exhibit a stronger shift between vehicle groups otherwise. From the distribution of vehicle groups along the regression line, it furthermore becomes evident that vehicles for which an accident was observed are more densely distributed toward the upper right corner of the graph, i.e. towards higher amounts of exposure, which is in accordance with expectations.

To extend our concept of exposure, we augment the aggregation of mileage and driving duration with information regarding the specific conditions under which the vehicle was operated. Firstly, we divide the 24 hours of a day into four intervals, which capture the different traffic conditions encountered during different times of the day such as morning commute, rush hour, evening, and late night. In each interval, both mileage and driving duration exposures are accumulated separately and are treated as input factors to the accident risk estimation problem. For the case of mileage aggregation, we consider two further types of variables are obtainable from location data: road type, as identified by urban, extra-urban, or highway, and the average velocity under which mileage was accumulated, divided into five intervals. These variables would not be available for the proposed ignition trigger-based exposure aggregation. Weekday as an additional variable was tested, but found non-significant with respect to the prediction of accident involvement for our sample.

We normalize the averaged, situation-related exposure variables to eliminate collinearity (Christensen 1997) for further modeling. Otherwise, these variables would vary according to the general level of exposure of an individual and effects would not be clearly separable. As a result of normalization, situation-related variables measure the fraction of exposure during which a vehicle was operated under respective conditions. For instance, a value of 0.1 for a certain time interval variable would signify that 10% of overall exposure (duration or mileage-based) were accumulated in this time interval. For descriptive statistics of absolute and situation-related exposure variables used, the reader is referred to Appendix B.

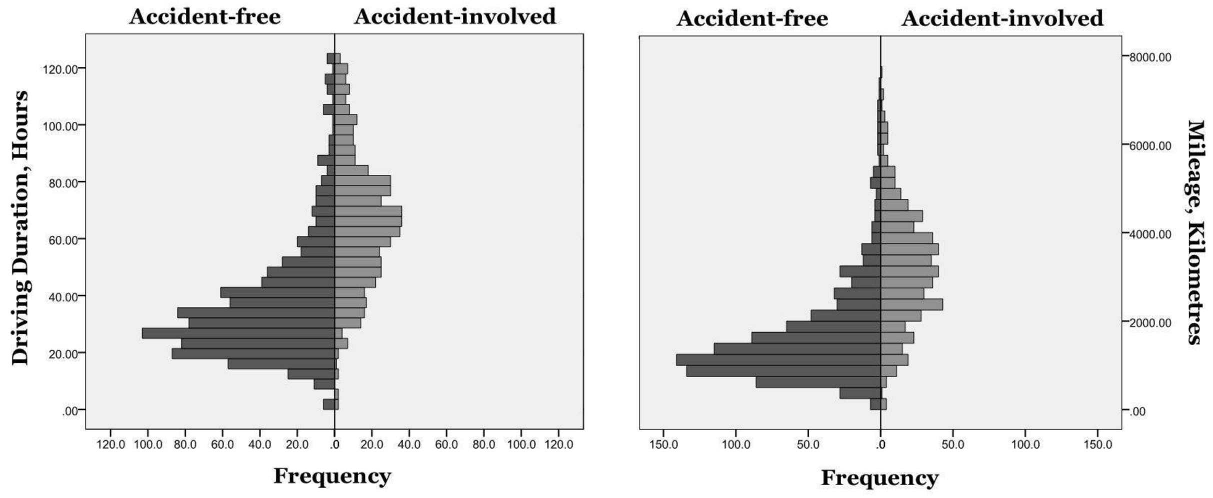

Figure 5. Exposure histograms for driving duration (left) and driven distance (right)

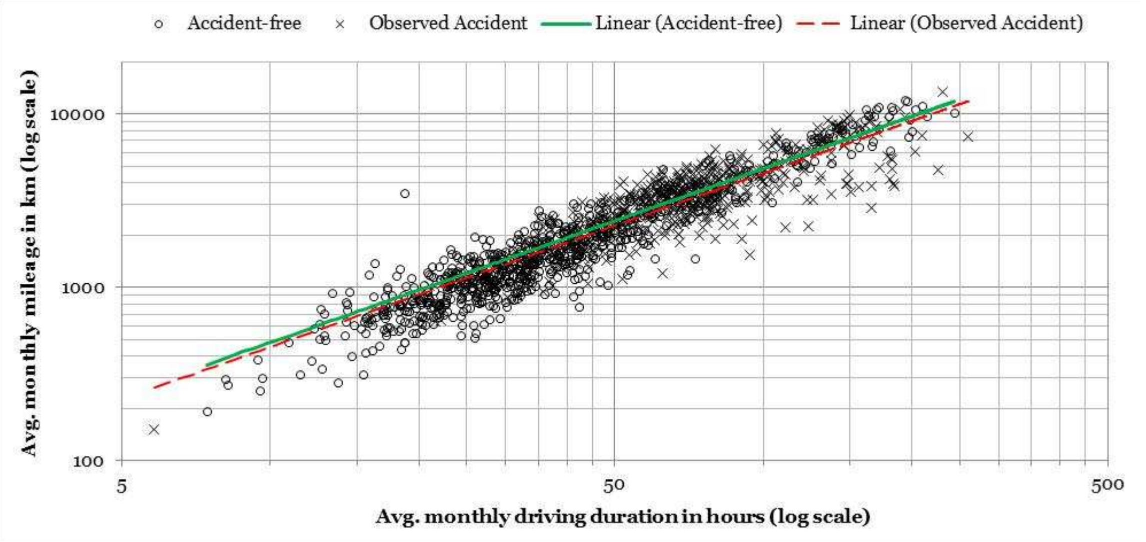

Figure 6. Vehicle driving duration versus mileage

Logistic Regression Model

In order to evaluate the different sources of exposure information, we estimate and compare a total of seven models. Two single-factor models are used to assess the predictability of risk of accident involvement from exposure alone. In these models, exposure is weighted with a single coefficient. Extended, multivariate models are used to incorporate the additional variables indicating the driving conditions under which exposure was accumulated. For the driving duration model, these comprise the four time-of-day intervals. With respect to mileage, further model variants include road type, velocity profile, or both, next to time-of-day. Their inclusion reflects the fact that a location-based UBI system would have access to these variables, while the proposed ignition-triggered design does not. The dependent variable in all seven models is the sample membership, i.e., accident-free (0) or accident-involved (1) vehicles. As an indicator for goodness of model fit, two common measures are Cox and Snell (1968) as well as Nagelkerke's (1991) pseudo-R^2 values, which are given for all four models in Table 1. Both measures determine the 'explained variance' of a specific model by comparing the prediction error of the full variable set with a model that includes only a constant predictor.

Table 1. Comparison of UBI design variants with different information types based on goodness of fit

Model	Cox and Snell's R^2	Nagelkerke's R^2
Driving Duration only	.206	.281
Driving Duration with time of driving	.363	.495
Mileage only	.212	.289
Mileage with time of driving	.321	.438
Mileage with time of driving, and road type	.367	.501
Mileage with time of driving, and velocity	.410	.560
Mileage with all variables	.416	.568

As expected, all models which include time-of-day information outperform the one factor models that use only a general exposure variable. Furthermore, model fit for the driving duration-based models exceeds the mileage-based variants. The highest achievable fit is .416 for Cox and Snell's and .568 for Nagelkerke's R^2, respectively. If road type and velocity are excluded, the duration-based model surpasses the mileage-based model. The additional gain in R^2 obtainable through their inclusion is .004/.006 for road type, .047/.065 for velocity, and .053/.073 for both combined. Velocity can hence be said to be significantly more important with respect to accident risk.

Variable coefficients and Wald significance values for the logistic regression models with base exposure and time interval variables are listed in Table 2. The time interval between midnight and 4:00 was excluded (stepwise forward method) as the inclusion of all four time-of-day variables resulted in high standard error terms. We attribute this observation to collinearity between the predictors, as the last time interval does not add information to the model. From the progression of B coefficients, it becomes evident that later time intervals are associated with a higher risk of accident involvement, which is in accordance with the literature.

Not displayed here, the additional situational variables in the mileage-based model exhibit the following characteristics. 'Urban' road type has a positive coefficient (B = 2.962), while it is negative for 'Highway' (B = -1.152). 'Extra-urban' does not add any additional information to the model and thus receives a coefficient equal to zero. The finding that driving within city limits is, per unit of exposure, riskier than on highways again agrees with established evidence. With respect to velocity, the interval between 60 and 90 km/h receives the highest lowest coefficient (B = -5.824), possibly in correlation to the road types on which these velocities typically occur.

Table 2. Coefficients (B) and p-values (Sig.) for base exposure and time interval variables

Variable	**Mileage-based model**		**Duration-based model**	
	B	Sig.	B	Sig.
Avg. monthly exposure (LN-transformed)	2.241	< .0001	3.243	< .0001
Time interval 5-18h	8.366	< .0001	12.175	< .0001
Time interval 18-21h	10.290	< .0001	18.458	< .0001
Time interval 21-24h	19.405	< .0001	22.070	< .0001

Discussion

Design Evaluation

Our evaluation has investigated the substitutability of location information-based UBI systems with less privacy-intrusive designs if certain situational risk factors are disregarded. A logistic regression analysis of both, distance and duration derived exposure metrics, yielded comparable predictive performance for both information types. This result challenges the prevalent notion of UBI as a location-based service and thereby may be an important design innovation to reduce PC toward UBI and improve its acceptance.

Road type and vehicle velocity – as the two major risk factors not accessible through an ignition-triggered UBI system – are in fact the two information types most critically viewed by consumers according to Figure 2. Also, particularly for road type the gain in predictive performance was negligible. Time of day, which on the other hand is accessible in the proposed design, is also indicative of accident risk and is only rejected by approximately a third of potential customers. If these figures are taken as proxies of technology acceptance, one can infer that by eliminating location and velocity from their UBI models, insurers significantly increase their potential market reach. This change in design would come with a minor reduction in the ability to estimate policyholders' accident risk, but is likely not to jeopardize the collection of the highly relevant exposure metrics. While we do not doubt that these considerations are somewhat optimistic, they do suggest that researchers and practitioners question established design paradigms of UBI.

The substitutability of location information has further implications on the realization of UBI systems from an insurance IT professional's perspective. Firstly, such a system is likely to be less technically complex, and less costly to implement. Moreover, it brings about fewer operational constraints as GPS sensors require an external antenna in line-of-sight of at least 3 satellites in order to determine valid position measurements and reduce the error in mileage aggregation. An ignition-trigger based solution may be mounted in concealed spots and thus does not interfere with the interior design of the car. Furthermore, GPS signal obstruction problems are circumvented, which reduces the likelihood of missing data and also of manipulation attempts where vehicle owners would block GPS reception to inhibit the recording of mileage. Lastly, GPS signal detection and processing requires significantly more power (Kjærgaard 2010), although this may not be the major concern for devices mounted in a regular road vehicle.

Limitations

We acknowledge several shortcomings in our analysis. Firstly, the evaluation sample contains vehicle-specific instead of driver or owner-specific data. It is not registered if one or several drivers are using a vehicle. However, we argue that this may also be seen as an advantage of UBI as driver variability is taken into account: While conventional insurance contracts usually based solely on the vehicle owner, UBI considers vehicle exposure independent of the driver that operates it. Secondly, the described dataset has limited external validity. It stems from a bounded geographic region and features a particular set of vehicles that may not be representative. In our opinion, this limitation does not devalue our results, as our contribution is mainly of conceptual nature and suggests that insurers implement the suggested design changes on their own discretion and based on additional empirical evidence.

Further limitations apply to our approach with respect to sample composition and data processing. The ratio of accident-free to accident-involved drivers in our sample is not justified by typical real-world accident frequencies in a driver sample. In consequence, our estimated model output does correlate to, but not correspond to actual accident probability. However, logistic regression with fixed intercept is a bias-free model that captures objective differences between groups regardless of prior distributions (Zadrozny 2004) and thus we see no reason for the reported predictive performance to deteriorate under varying sample compositions.

Besides these technical issues, our results are also limited in that they only cover an isolated aspect of PC. We presume that the substitution of location data – which our results suggests to be feasible without compromising system performance – will in itself result in an increased acceptance of UBI among consumers. This premise remains hypothetical, and it is conceivable that in spite of the evidence presented in Figure 2 the improved design will still face rejection in the market. In particular, PC may arise not only from the collected vehicle data per se, but from the larger context in which information regarding individual driving behavior is distributed and used. While such considerations are outside the scope of the analysis presented in this paper and do not affect the validity of reported results, it is imperative that they be investigated in subsequent work on UBI system privacy.

Suggestions for Further Research

This paper has demonstrated the importance of a holistic, socio-technical IS perspective in privacy research. Further studies are required that elucidate the interaction between system design, actuarial performance and privacy perceptions in UBI. As a relatively novel application that couple the massive collection and evaluation of ubiquitous data with insurance services, a field that demands high trust and long-term relationships between customers and vendors, UBI are a particularly interesting domain of privacy-related IS research. While our research has taken a design science approach to address conceptual issues in UBI system design, we call for behavioral and psychological perspectives to complement our results. Furthermore, it remains for more technically inclined scholars to implement the proposed design of UBI without location information to proof its feasibility in the field, ideally in collaboration with an insurance provider.

In a broader view, there appear to be several unanswered question at the intersection of ubiquitous IS design and privacy research. For most ubiquitous computing applications, designers face several options with respect to the chosen data collection methods. While location information has become pervasive and is usually very effective, it is allegedly also the most privacy-intrusive type of information and may therefore largely determine the acceptance of a design. As a tool for practical decision making, researchers should extend current models and constructs of IS privacy to incorporate the influence of varying information types, individually and in combinations. Specifically the emerging field of privacy calculus may benefit from such a differentiated approach and contribute substantially to the success of ubiquitous systems in the future.

Besides consumers' concerns and perceptions, privacy should also be subjected to a more rigorous treatment in terms of the efforts and benefits its provision entails for business stakeholders of ubiquitous IS. The pursued design science research approach has highlighted the importance of an evaluation of privacy that takes functional business requirements into account. Through usefulness and related constructs in technology acceptance research it is also evident that the fulfillment of these requirements may be reflected in adoption success just as much as the safeguarding of privacy. Finally, we call for the research community to develop design frameworks that allow for a holistic analysis of information type variants in ubiquitous IS that integrate privacy assessments from both, consumers and system operators. Such an extended concept of privacy calculus may find broad adoption in application areas such as UBI, road pricing, and car sharing in the context of transportation, and in more general contexts such as digital commerce and counter-crime surveillance.

Conclusion

This paper has pursued a design science research approach to privacy enhancement of UBI. Our research was motivated by articulate differences in privacy perceptions among consumers regarding different types of information. We proposed a novel method of collection of general and situation-specific exposure data that makes location information obsolete in UBI. In order to evaluate this design artifact, we emulated its information base from a sample of high resolution GPS trajectories. A comparative assessment of predictive power between the proposed artifact and conventional UBI design yielded the equivalence of location information and ignition-trigger derived exposure as factors in UBI ratemaking.

We inductively extrapolated our results to the more general domain of ubiquitous IS privacy, outlining the necessity of a systemic, socio-technical perspective in ubiquitous IS design that takes into account the privacy concerns and business value of different types of information. We listed several entry points for further research in this direction and call for an extension of privacy calculus concepts to include information type. Following the proverb “a bird in the hand is worth two in the bush”, there are substantial grounds to demand that ubiquitous IS design follows the familiar Occam's Razor Principle in the selection of information to be collected, instead of exhausting the technically feasible.

Our research also entails several implications for professionals from insurance, component supplier and telematics service provider industries, as well as for regulatory authorities. Dominant designs in state of the art UBI systems should be challenged and, if proven successful in further studies, designs that omit location information be implemented. As the analysis presented in this paper has demonstrated, high resolution GPS measurements can be an overkill and might jeopardize privacy if vehicle exposure and situational factors are all that is required for premium calculation. Systems designers should carefully balance privacy-intrusiveness of their systems against added value from more comprehensive services. More moderate designs may gain increased trust and acceptance and allow business stakeholders and consumers alike to reap the ample benefits associated with UBI.

Appendix A. Design Characteristics of UBI Systems in Europe

Based on Troncoso et al. (2007), market statistics (Frost and Sullivan 2010), and own research. Includes available offerings as well as pilots, provider list may not be comprehensive. Where location of driving – i.e., usually road type – is not an explicit rate factor, it is used to determine vehicle mileage.

Table A. Design Characteristics of UBI Systems in Europe

Insurance Provider (Product Name)	Country	Published Rate Factors	Uses Location Information
Uniqua (SafeLine)	Austria	Mileage	Yes
Axa G-Box	Belgium	Acceleration	No
P&V (Go Box)	Belgium	Acceleration	No
Vivium (S2)	Belgium	n.a.	Yes
April	France	n.a.	Yes
Aviva (Avantage Kilometres)	France	Mileage	Yes
Axa (PAYD 4000)	France	Mileage	Yes
Groupama (Amaguiz)	France	n.a.	Yes
Solly Azar (Easy Drive)	France	Mileage, time of day	Yes
Aurora Assicurazioni (Aurobox)	Italy	n.a.	Yes
Fondiaria	Italy	n.a.	Yes
Generali (Auto GPS)	Italy	n.a.	Yes
Gruppo ITAS Assicurazioni (Fido)	Italy	n.a.	Yes
INA Assitalia (OttoSat)	Italy	n.a.	Yes
Italiana Assicurazioni	Italy	n.a.	Yes
Linear Assicurazioni (Linearsat)	Italy	Mileage, location	Yes
Navale Assicurazioni (Navalbox)	Italy	n.a.	Yes
Pacifica	Italy	n.a.	Yes
Reale Mutua (Full Box)	Italy	n.a.	Yes
Sara (Sarafree)	Italy	Mileage	Yes
Unipol Assicurazioni (Unibox)	Italy	Mileage, location	Yes
Allianz	UK	Mileage, time of day, location	Yes
Equity Red Star	UK	Mileage, time of day, location	Yes
Groupama	UK	Mileage, time of day, location	Yes
Insure The Box	UK	Mileage, Acceleration	Yes
The Co-Operative	UK	Mileage, time of day, location	Yes
Mapfre (Y CAR)	Spain	Mileage, time of day, velocity, location	Yes

Appendix B. Descriptive Statistics of Explanatory Variables

Separate according to subsample (accident-free, accident-involved).

Table B1. Descriptive Statistics of Mileage-based Variables

Variable	**Accident-free**		**Accident-involved**	
	Mean	Std. Dev.	Mean	Std. Dev.
Avg. monthly exposure (LN-trans.)	7.3159	.82695	8.0749	.54859
Time interval 0-5h	.0798	.12407	.0612	.07678
Time interval 5-18h	.7879	.12120	.7585	.15190
Time interval 18-21h	.0929	.07005	.1173	.07188
Time interval 21-24h	.0393	.04244	.0631	.06599
Road type (urban)	.3195	.17915	.3135	.18019
Road type (extra-urban)	.3198	.21452	.3150	.16782
Road type (highway)	.3608	.17877	.3716	.20297
Velocity interval < 30 km/h	.1714	.10429	.1373	.08166
Velocity interval 30-60 km/h	.2650	.09862	.2465	.10755
Velocity interval 60-90 km/h	.3153	.17606	.2747	.15944
Velocity interval 90-120 km/h	.1417	.09760	.1821	.10013
Velocity interval >120 km/h	.1067	.12749	.1593	.14270

Table B2. Descriptive Statistics of Duration-based Variables

Variable	**Accident-free**		**Accident-involved**	
	Mean	Std. Dev.	Mean	Std. Dev.
Avg. monthly exposure (LN-trans.)	3.7263	.60892	4.3194	.44942
Time interval 0-5h	.0694	.11188	.0512	.06535
Time interval 5-18h	.8017	.11143	.7739	.14032
Time interval 18-21h	.0916	.06566	.1154	.06795
Time interval 21-24h	.0372	.03967	.0595	.06152

References

Angst, C. M., and Agarwal, R. 2009. "Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likehlihood Model and Individual Persuasion," *MIS Quarterly* (33:2), pp. 339-370.

Belanger, F., and Crossler, R. E. 2011. "Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems," *MIS Quarterly* (35:4), pp. 1017-1041.

Bolderdijk, J. W., Knockaert, J., Steg, E. M., and Verhoef, E. T. 2011. "Effects of Pay-As-You-Drive Vehicle Insurance on Young Drivers' Speed Choice: Results of a Dutch Field Experiment.," *Accident Analysis and Prevention* (43:3), pp. 1181-6.

Buzzacchi, L., and Valletti, T. M. 2005. "Strategic Price Discrimination in Compulsory Insurance Markets," *The Geneva Risk and Insurance Review* (30:1), pp. 71-97.

Chiappori, P.-A., Jullien, B., Salanié, B., and Salanié, F. 2006. "Asymmetric Information in Insurance: General Testable Implications," *RAND Journal of Economics* (37:4), pp. 783-798.

Christensen, R. 1997. *Log-linear Models and Logistic Regression*, 2nd ed., New York: Springer.

Coroama, V. 2006. "The Smart Tachograph - Individual Accounting of Traffic Costs and its Implications," in *Proceedings of the 4th Int. Conference on Pervasive Computing*, Dublin, Ireland, pp. 135–152.

Cox, D. R., and Snell, E. J. 1968. "A General Definition of Residuals," *Journal of the Royal Statistical Society. Series B (Methodological)* (30:2), pp. 248–275.

Dinev, T., and Hart, P. 2006. "An Extended Privacy Calculus Model for E-Commerce Transactions," *Information Systems Research* (17:1), pp. 61-80.

Duri, S., Gruteser, M., Liu, X., Moskowitz, P., Perez, R., Singh, M., and Tang, J.-M. 2002. "Framework for Security and Privacy in Automotive Telematics," in *ACM Proceedings of the 2nd International Workshop on Mobile Commerce*, Atlanta GA, USA, pp. 25-32.

Fano, A., and Gershman, A. 2002. "The Future of Business Services in the Age of Ubiquitous Computing," *Communications of the ACM* (45:12), pp. 83-87.

Filipova-Neumann, L., and Welzel, P. 2010. "Reducing Asymmetric Information in Insurance Markets: Cars with Black Boxes," *Telematics and Informatics* (27:4), pp. 394-403.

Fleisch, E., and Thiesse, F. 2007. "On the Management Implications of Ubiquitous Computing: An IS Perspective," in *ECIS 2007 Proceedings*, St. Gallen, Switzerland, pp. 1929–40.

Frees, E. W. 2008. "Hierarchical Insurance Claims Modeling," *Journal of the American Statistical Association* (103:484), pp. 1457-1469.

Frost and Sullivan. 2010. *Strategic Analysis of the European Market for Telematics-enabled Usage-based Insurance*. Report.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-105.

Hong, S.-J., Thong, J.Y.L., Moon, J.-Y., Tam, K.-Y. 2008. "Understanding the Behavior of Mobile Data Services Consumers," *Information Systems Frontiers* (10:4), pp. 431-445.

Iachello, G., and Hong, J. 2007. "End-User Privacy in Human-Computer Interaction," *Foundations and Trends in Human-Computer Interaction* (1:1), pp. 1-137.

Iqbal, M. U., and Lim, S. 2010. "Privacy Implications of Automated GPS Tracking and Profiling," *IEEE Technology and Society* (Summer), pp. 39-46.

Jun, J., Guensler, R., and Ogle, J. 2010. "Differences in Observed Speed Patterns Between Crash-involved and Crash-not-involved Drivers: Application of In-vehicle Monitoring Technology," *Transportation Research Part C: Emerging Technologies* (19:4), pp. 569-578.

Langheinrich, M. 2001. "Privacy by Design – Principles of Privacy-Aware Ubiquitous Systems," in *Ubicomp 2001: Ubiquitous Computing, Lecture Notes in Computer Science* (2201), pp. 273-291.

Kohli, R., and Grover, V. 2008. "Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times," *Journal of the Association for Information Systems* (9:1), pp. 23-39.

Litman, T. 1997. "Distance-based vehicle insurance as a TDM strategy," *Transportation Quarterly* (51:3), pp. 119-138.

Mayer, P. 2012. "Empirical Investigations on User Perception and the Effectiveness of Persuasive Technologies," *Manuscript submitted for publication*.

Mayou, R., Bryant, B., and Duthie, R. 1993. "Psychiatric Consequences of Road Traffic Accidents," *British Medical Journal* (307:6905), pp. 647.

McParland, C., and Connolloy, R. 2008. “Empirical Research on Technology-Related Privacy Concerns: A Review and Critical Assessment,” in *ICIS 2008 Proceedings*.

Nagelkerke, N. J. D. 1991. “A Note on a General Definition of the Coefficient of Determination,” *Biometrika* (78:3)Biometrika Trust, pp. 691–692.

Paefgen, J., Michahelles, F., and Staake, T. 2011. "GPS trajectory feature extraction for driver risk profiling," *International Workshop on Trajectory Data Mining and Analysis at the 13th Conference on Ubiquitous Computing*, Beijing, China.

Park, I. 2009. “The Study on The Relationship Between Privacy Concerns and Information Systems Effectiveness,” in *ICIS 2009 Proceedings,* Phoenix, Arizona.

Pee, L. G. 2011. “Attenuating Perceived Privacy Risk of Location-Based Mobile Services,” in *ECIS 2011 Proceedings*.

Peffers, K., Tuunanen, T., Rothenberger, M., and Chatterjee, S. 2007. “A Design Science Research Methodology for Information Systems Research,” *Journal of Management Information Systems* (24:3), pp. 45-77.

Phelps, J., Nowak, G., and Ferrell, E. 2000. “Privacy Concerns and Consumer Willigness to Provide Personal Information,” *Journal of Public Policy and Marketing* (19:1), pp. 27-41.

Popa, R. A., Balakrishnan, H., and Blumberg, A. J. 2009. “VPriv: Protecting Privacy in Location-Based Vehicular Services,” in *Proceedings of the 18th Conference on USENIX Security Symposium*Berkeley CA, USA.

Prince, K., and Barrett, M. 2005. “Privacy Implications of Technology Innovation Processes,” in *ICIS 2005 Proceedings*, pp. 423-433.

Progressive Insurance. 2005. *Texas Mileage Study□: Relationship Between Annual Mileage and Insurance Losses*. Report.

Sen, S., Raghu, T. S., and Vinze, A. 2009. “Demand Heterogeneity in IT Infrastructure Services: Modeling and Evaluation of a Dynamic Approach to Defining Service Levels,” *Information Systems Research* (20:2), pp. 258-276.

Sheng, H., Nah, F. F.-H., and Keng, S. 2008. “An Experimental Study on Ubiquitous Commerce Adoption: Impact of Personalization and Privacy Concerns,” *Journal of the Association for Information Systems* (9:6), pp. 344-376.

Spiekermann, S., and Cranor, L.F. 2009. “Engineering Privacy,” *IEEE Transactions on Software Engineering* (35:1), pp. 67-82.

Thiesse, F. 2007. “RFID, privacy and the perception of risk: A strategic framework,” *The Journal of Strategic Information Systems* (16:2), pp. 214-232.

Toledo, T., Musicant, O., and Lotan, T. 2008. “In-vehicle Data Recorders for Monitoring and Feedback on Drivers’ Behavior,” *Transportation Research Part C: Emerging Technologies* (16:3), pp. 320-331.

Troncoso, C., Danezis, G., Kosta, E., and Preneel, B. 2007. “PriPAYD: Privacy Friendly Pay-As-You-Drive Insurance,” In *Proceedings of the 2007 ACM Workshop on Privacy,* Alexandria VI, USA.

Xu, H., Dinev, T., Smith, H. J., and Hart, P. 2008. “Examining the Formation of Individual’s Privacy Concerns: Toward an Integrative View,” in *ICIS 2008 Proceedings*.

Xu, H., and Teo, H.-H. 2004. “Alleviating Consumers’ Privacy Concerns in Location-Based Services: A Psychological Control Perspective,” in *ICIS 2004 Proceedings*, pp. 793-806.

Xu, H., and Teo, H.-H. 2005. “Consumers’ Privacy Concerns Toward Using Location Based Services: An Exploratory Framework and Research Proposal,” in *ICIS 2005 Proceedings*.

Xu, H., Teo, H.-H., and Tan, B. 2005. “Predicting the Adoption of Location-Based Services: The Role of Trust and Perceived Privacy Risk,” in *ICIS 2005 Proceedings*, pp. 897-910.

Xu, H., Teo, H.-H., Tan, B. C. Y., and Agarwal, R. 2009. “The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services,” *Journal of Management Information Systems* (26:3), pp. 135-174.

Yun, H., Han, D., and Lee, C. 2011. “Extending UTAUT to Predict the Use of Location-Based Services,” in *ICIS 2011 Proceedings*.

Zadrozny, B. 2004. “Learning and Evaluating Classifiers Under Sample Selection Bias,” in *ACM Proceedings of the Twenty-first International Conference on Machine Learning,* Banff, Canada.

Zhou, T. 2011. “The Impact of Privacy Concern on User Adoption of Location-Based Services,” *Industrial Management and Data Systems* (111:2), pp. 212-226.

Zuo, Y. 2010. “A Holistic Approach for Security Requirement Specification for Low-cost, Distributed Ubiquitous Systems,” in *ICIS 2010 Proceedings,* St. Louis, USA.