E-privacy: Evaluating a new search cost in online environments

Dirk Annacker

Humboldt University of Berlin
Institute of Marketing
Spandauer Straße 1
10781 Berlin
Germany
Tel: +493020935751
Fax:+493020935675
annacker@wiwi.hu-berlin.de

Sarah Spiekermann

Humboldt University of Berlin
Institute of Information Systems
Spandauer Straße 1
10781 Berlin
Germany
Tel:+49308860482
sspiek@wiwi.hu-berlin.de

Martin Strobel

International Institute of Infonomics
PO Box: 26066401 DC
6412 Heerlen
The Netherlands
Tel: +21455707686
martin.strobel@infonomics.nl

Dirk Annacker, Sarah Spiekermann, Martin Strobel

ABSTRACT:

Electronic Commerce environments increasingly witness a conflict on the subject of e-privacy: While marketers want to maximize their customer knowledge and grasp the identity of their online users, consumers often want to stay anonymous and not reveal private information. The conflict suggests that 'private consumer information' should be respected as a new search cost for consumers in EC environments. The current paper aims to 'grasp' the phenomenon of this new search cost entitled as 'private consumer information cost' (PCIC). The paper aims to evaluate PCIC by identifying its main drivers and their interrelation. An empirical study is presented which shows that three factors, perceived importance, legitimacy and difficulty of online requests made by marketers in a purchase context explain much of the variance of PCIC. Empirical data also reveals how different types of information requests drive PCIC. The types of information distinguished are product information, information on product usage as well as personal information. Results hint at the fact that consumers accept personal information requests to a greater extend than one would expect, but only as long as they improve product- or service choice. It is concluded that marketers incur considerable opportunity cost of information if they do not respect the nuances evident in PCIC creation and do not rely on them for the strategic design of their online communication.

1. Introduction

Understanding consumers' information search behavior prior to the purchase of goods is critical to firms' strategic marketing activity. It therefore has a long tradition in economics as well as marketing theory [4,7,18,22,26,27]. Traditional search cost analysis for offline markets has focused on consumers' physical effort to compare products, travel expenditure, information processing- and time cost. With the advent of the Internet these traditional search cost incurred by consumers have to be reinvestigated: EC sites, and especially infomediaries, help consumers save time and effort when they search for products and facilitate the complex combination and comparison of goods through the use software agents (recommendation- and search engines).

However, new search cost factors may also be created by the use of the Internet. One new cost factor relevant in online environments seems to be the cost of privacy. This is due to the particularly threatening capability of the electronic medium to link user data and to create customer profiles [3, 28]. While customer information has increasingly been recognized as an important asset for companies that drives competitive advantage [13,23], many consumer surveys show that online users are afraid of losing their privacy online [1,21,29]. Their fear often expresses itself in service denial, or, even more often, in the provision of false personal data [3,9].

On this background, we want to introduce the idea that online consumers are confronting a new dimension of search cost on the Internet which we call 'private consumer information cost' (PCIC). Consumers experience this cost when revealing 'truthful' information about themselves on the Internet while knowing that afterwards some parts of their identity will be known to the organization hosting a site and that their data will probably be used for further analysis or for sale.

We claim that if marketers respected information provision as an online search cost to their users they would probably pay more attention to offer appropriate benefits in return for private data[1]. In fact, studies have revealed that people are ready to reveal information, but only if they receive appropriate returns [13]. As a result, marketers have to learn how people evaluate their data and consequently their electronic privacy. They have to win a feeling for *what* and *how much* they can actually ask online.

What has been missing from research up to now, however, is an insight into the ways in which people 'evaluate' their private data. Hine and Eve stated in 1998 [13]: *"Despite the wide range of interests in privacy as a topic, we have little idea of the ways in which people in their ordinary lives conceive of privacy and their reactions to the collection and use of personal information."* Unfortunately, studies that aimed to explore the phenomenon since then only focused on the provision of single data units (such as the provision of an e-mail address), but

[1] It has been recognized by scholars as well as research institutes that appropriate returns are vital to online success[12]. Yet, still there are many EC sites which ask users to fill out electronic questionnaires where the benefits for the person answering are not obvious. There are also product search engines online that ask consumers to specify every detail of the desired product, but are not able to provide a satisfying recommendation in return. Frequently, online users are asked to provide information about their location, age and reaching data, but this data demand has nothing to do with the context for which users visited the site and it is unclear why they should provide it.

never reflected on the context in which information units could be requested on the Internet [1]. With this, they failed to respect the importance of the *context* for information valuation that has been recognized for long in the information science literature [2,14].[2]

Seeing this gap in research we developed a simple model with the goal to reflect a user's context-related cost of online information provision. The challenge we confronted in developing the model is that no tangible value is actually capable of representing PCIC appropriately. There is usually no cost created to produce private information. Economic freebies or services so far offered in exchange for PCIC strongly differ in value [9]. Also, what is regarded as a 'high-cost' information by one individual is perceived as 'low cost' by another. It is therefore not possible to attribute a specific value unit (e.g. a monetary unit) to one specific information unit that would be acceptable for everybody. The model presented hereafter therefore focuses more on the identification of some overall variables driving PCIC and their interrelation. In section 2 we will present the variables we identified to be important for PCIC and how we derived them. In section 3 we will present an empirical study we carried out in order to test the three variables' impact on PCIC, their interrelation and practical implication for communication design in EC web sites. In section 4 results of the empirical study will be discussed and some practical advice will be deducted for the design of communication between interface agents and consumers. Section 5 summarizes the major findings that can be deducted from the PCIC model and includes some propositions for future research.

2. Identifying Relevant Drivers of Personal Information Cost on the Internet

When people provide information about themselves on web sites they usually do so either by 'chatting' freely (e.g. in communities) or by answering concrete questions (e.g. product configuration engines, online questionnaires, etc.). For the purpose of this article we are focusing only on the evaluation of PCIC for the latter context, because we believe this type of online communication to play an important role in Electronic Commerce[3].

Personal consumer information cost in the way we define it stands for the loss in utility a consumer perceives when giving away a truthful information unit about himself. PCIC expresses itself in a consumer's reluctance to answer the question of an interface agent in the context of a product search process. Strong reluctance stands for high information cost. In contrast, if a user has no problem to reveal an information unit about himself he incurs little cost.

As the determination of PCIC means to attribute value to different types of information units, research in information theory provides a starting point for modeling. Considerable research has been done on the valuation of information in management science (particularly decision theory) as well as in the humanities. None of these approaches are directly transferable to the current context, but some principal theoretical constructs of information valuation can still be applied[4]; notably the influence of the *context* on information value, the *relevance* the information unit holds in this context and the *effort* required to *process* it [2].

The context in which an information unit is demanded can influence the perception of PCIC. As Badenoch et al. [2] resume, the „*value [of information] is almost entirely dependent on the specific circumstances in which the information will be used"*. A practical example may illustrate this: Let's assume a buyer who wants his goods to be delivered to the home. He will probably be most open to provide his address to the supplier. The delivery context creates the necessity to provide the address and thus legitimizes its provision. If, in contrast, the customer picked up the ordered products himself, he would probably be surprised if he had to leave his address with the vendor for there is no obvious contextual need for this information provision. It is likely that he would be reluctant to provide it. The example shows that the *perceived legitimacy* of an information requested in a specific context drives the perceived cost of providing it. As Hine and Eve put it [14]: *"Requests for information not deemed necessary in order to carry out this function were deemed intrusive."* The arguments suggest that the perceived legitimacy of a question influences PCIC. It therefore represents one dimension in the PCIC evaluation model presented hereafter. It is defined as the degree to which a question is perceived as justified in a given context.

[2] For example, in one context users might perceive the provision of their telephone number as a necessity and are therefore most willing to give it away (no/little cost). In other contexts, they might regard the provision of the telephone number as an unnecessary intrusion into their privacy and will only reluctantly provide it (high cost). In this latter situation, a marketer would be well advised to explicitly offer annoyed users some tangible returns for their input.

The legitimacy of an information request is not only determined by the context, but also by its *importance* in that context. In the above example, providing the delivery address is very important for the fulfillment of the service. It is therefore intuitive to argue that the buyer perceives little cost to provide it. Yet, there may be other legitimate information units in the delivery context which are less important and thus are perceived more costly to provide. For example, the telephone number of the product recipient or his working hours. The *perceived importance* of an information unit in a specific context thus also has a strong impact on the perception of PCIC and the subsequent willingness to provide it. For modeling purposes we define importance as the perceived degree to which an information request can contribute to an optimal product or service experience.

While importance drives the legitimacy of an information request, the opposite does not hold true. For example, asking the buyer of a winter jacket what type and color of buttons he prefers may be a legitimate question in the purchase context, but will probably not be important to most consumers.

Finally, it has been recognized in literature that the effort to process information also leads to cost for consumers [7]. Eventually, there may be information requests online that are difficult for users to answer. As a result, they may be reluctant to do so. For example, if a search engine asked for the envisaged gigabyte size of a hard disc, but the user does not know what a hard disc is. The *perceived difficulty* to answer a question represents the third dimension of the PCIC evaluation model.

The three main drivers of PCIC, identified as perceived legitimacy, importance and difficulty to provide an information unit in a specific context are summarized in Figure 2. They are at the core of the empirical investigations presented hereafter to better understand the construct of PCIC. Certainly, they are not able to explain the phenomenon of PCIC in its entirety. Individual differences, for example, in the individual level of trust in online providers, online privacy attitudes, product experience etc. may also drive the level of PCIC. Yet, as will be shown below, the three variables examined represent a good starting point for the understanding of PCIC and strategic marketing responses to it.

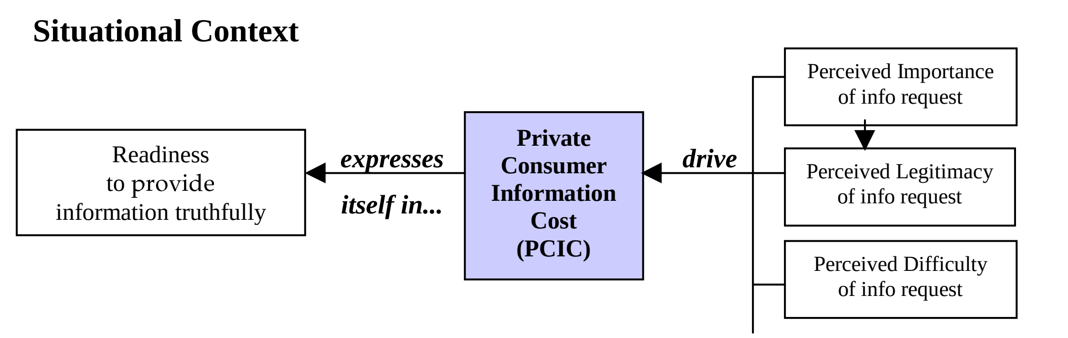

Figure 2: Drivers of Personal Consumer Information Cost (PCIC)

3. Experimental Design and Data Collection

Selling high-involvement goods over the Internet implies a detailed question-answer process between interface agents and consumers. To design this process it is important to know what questions can be asked by the interface agent and how they have to be formulated in order to minimize PCIC. So far, EC Web sites usually restrict their communication to an exchange of preferences for different product attributes. Very few personal or usage-related questions are asked [25] and mostly web design is focused on a minimization of time-cost for consumers.

The goal of the empirical study was therefore to examine how the request of different information units, also highly personal and usage oriented ones, drives consumers' perception of PCIC and how the three dimensions introduced in section 2 contribute to this. 39 subjects were invited to the university laboratory at Humboldt University Berlin and were asked to judge 112 questions that could potentially be asked by an electronic sales agent in a WWW store. 56 questions displayed for judgment to the subjects were linked to the purchase of a winter jacket. The following 56 questions could be asked during the selection of a compact camera Even though one could argue that asking consumers 56 questions online in a sales context is rather unrealistic other experimental studies we conducted suggest that this is not the case [24].[3]

All questions were initially developed with the help of 'real-world' sales agents selling these two product categories in a premium department store. All questions were linked to the product choice, but not to payment or delivery issues. They were formulated in such a way that they would all directly address the user in person (e.g. Do *you*….) and be of multiple choice nature. They were purposefully developed to represent four distinct content categories: 1) questions addressing product attributes (*pd*) (e.g.: How resistant do you want the fabric of *the* jacket to be?), 2) those looking into the usage envisaged with the product (*u*) (e.g.: Where do you want to wear the jacket?), 3) personal questions completely independent of the product (*peip*) and 4) personal questions related to the product (*pepr*). While *peip*-questions are linked to the communication context they have no influence on the selection algorithm (e.g.: Where do you obtain your knowledge about fashion?). In contrast, *pepr*-questions do support the search process, but also capture a lot of information on a person's general view on the respective product category (e.g.: How important is the resistance of the fabric of *jackets* to you?). Table 1 gives some concrete examples for the four question categories (in the real questionnaire typically 4-6 possible answers are provided).

Product	Q-Type	Q-Text	Q-Answer 1	Q-Answer 2
Camera	*pd*	How strong do you want the zoom of the camera to be?	140-170 mm	101 -139mm
Jacket	*pd*	What size do you need for the jacket?	XS	S
Camera	*u*	At what occasions do you usually take photos?	Vacation	Parties
Jacket	*u*	At what occasions do you want to wear the jacket?	at the office	at the client
Camera	*pepr*	How important are to you relatively cheap photo development cost?	very important	important
Jacket	*pepr*	How important are to you the recognition of trend models?	very important	important
Camera	*peip*	What is your motivation when taking photographs?	Fun	Arts
Jacket	*peip*	How often do you buy a new jacket?	very often: > 2 times per seas.	often: every season

Table 1: Examples for Different Question Types and Potential Answers

Interface questions and potential multiple choice answers were displayed to subjects one after another on the left side of a computer screen. Subjects were asked to imagine that the questions displayed to them would be asked by a product search engine on the Internet in the context of a purchase process. On the right side of the screen an 11 point scale (from 0 to 10) simultaneously asked subjects to judge each question's legitimacy and importance in the sales context, the difficulty to answer it as well as the overall perceived information cost. The construct of information cost was explained to the participants in advance of the rating sessions through a text based briefing which used the following definition of PCIC: *Information Cost is standing here for the 'intuitive readiness' to truthfully answer the question of the search engine; thus the spontaneous feeling, whether you would be willing to reveal the demanded information about yourself. 'No' Information Cost would mean that you have no problem at all to answer the question truthfully. 'Very high' Information Cost stands for the emotion that under no circumstances you would give this type of information about yourself to a search engine.*

[3] In the context of the IWA experiments at Humboldt-Universität we in fact discovered that online users enjoy rather deep and personal communication features online if they search for high-involvement goods. In these experiments an anthropomorphic 3-D shopping bot was used to ask potential online buyers precisely those 56 questions that we comment on in this paper. It turned out that 54% of shoppers answered at least 98% of questions displayed to them [24]. For more detail on the experiments see: http://iwa.wiwi.hu-berlin.de

4. A Model for PCIC

For modeling purposes one outlier had to be excluded from the initial number of 39 observations. The model presented hereafter is therefore based on 38 observations.

4.1. Initial Regression Analysis

The relationship between information cost (*PCIC)* as the dependent variable and legitimacy (*Leg*), importance (*Imp*), and difficulty (*Diff*) as independent variables can be expressed as:

$$PCIC_{ij} = \beta_0 + \beta_1 Leg_{ij} + \beta_2 Imp_{ij} + \beta_3 Diff_{ij} + \varepsilon_{ij}, \tag{1}$$

where: $i = 1, \ldots, I$ number of respondents, $j = 1, \ldots, J$ number of questions.

As ordinary least square analysis of this model (1) resulted in a relatively low R^2 of .439 for pooled data, $F(3, 4252) = 1108.69$, $p < .01$, we estimated an alternative model where unobserved heterogeneity was captured by dummy variables for each respondent (Table 2).

Overall model fit		
$R^2 = .623$ Adj. $R^2 = .619$ $F(40, 4215) = 173.80$, $p < .01$		
Parameter estimates		
Independent variables	Parameter	Dependant variable: *PCIC*
Intercept	β_0	6.252
Leg	β_1	-.559 (.017) ***
Imp	β_2	-.011 (.018)
Diff	β_3	.138 (.014) ***
( ) standard error; *** $p < .01$ Since the data consists of partially dependent observations, controlling for these dependencies might lead to slightly lower levels of significance.		

Table 2: Model Results for the Fixed Effects Regression Model

As can be seen from Table 2, model (1) has an acceptable fit. The signs of all parameters support the expectation that legitimacy and importance lead to a reduction in PCIC while the difficulty of an information request influences it positively. Surprisingly, however, the impact of perceived question importance does not appear significant. One reason for this result may be the bivariate correlation of .825 between *Leg* and *Imp.* Co-linearity diagnostics shows that the largest condition index (18.50) is above 15 which, according to Belsley et al. [5], indicates a borderline case of co-linearity.

As co-linearity problems subsequently lead to ambiguity in interpretation of results, we decided to explore the relationship between *Leg* and *Imp* in more detail (Figure 3).

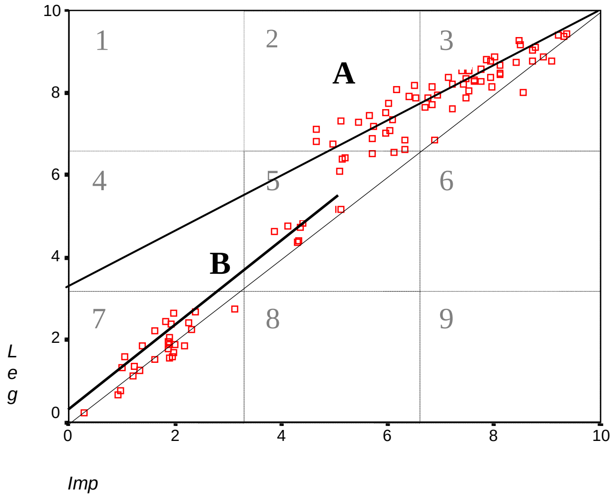

Figure 3: Scatter Plot for Mean Values of Leg and Imp

4.2. Coherences between Legitimacy and Importance of Information Requests

In order to allow for better interpretation of the data, the relationship between legitimacy and importance was moved from the disaggregated level to the aggregated level. Here, we computed the mean values of both variables for all questions. Figure 3 gives an overview of the observations made. Besides a strongly apparent linear relationship between legitimacy and importance of interface questions it is interesting to note that questions can apparently be separated into two distinct groups: For questions in the lower left corner (represented by graph B) an increase of one scale point in importance seems to correspond to a similar increase in legitimacy. In contrast, for questions in the upper right corner the increase in legitimacy is noticeably smaller (graph A).

In order to analyze the nature of these two apparently distinct relationships, we included the nature of questions into our interpretations. As was discussed in section 3 questions were purposefully designed to represent four different content categories: Questions could either be related to the product (*pd*) or its usage (*u*). They could address personal traits only (*peip*) or ask for a more general view of the person on a product category (*pepr*). Transferring this typology to the two distinct graphs (A and B), it is interesting to note that group A of questions (represented by graph A) are primarily product related questions (*pd*) as well as person oriented questions with a product focus (*pepr*). At the same time, group B (represented by graph B) are mostly questions focusing on personal attributes (*peip*) or usage (*u*).

To go into more detail, we divided both scales into three sections (0 – 3.33, 3.34 – 6.66, 6.67 – 10) and created 9 different classes for Leg x Imp. As can be seen in Figure 3 there are only 5 classes relevant to the analysis (classif1): class 7 containing questions of low legitimacy and importance, classes 2 and 3 containing in contrast highly legitimate and important questions and class 5 where legitimacy and importance are medium. Class 4 which only contains two items appears negligible for future discussion. Table 3 gives an overview of those types of questions that are present in the different classes. We are aware of the scientific restrictions of table 3 as some of the cross-tabulation categories contain a very small number of observations. However, we still feel that the discussion of the table provides some valuable insights and hints for future research on this subject.

As would be expected, more than 95% of product attribute questions (*pd*) were perceived as highly legitimate by subjects while over 80% of solely person oriented questions (*peip*) were perceived as little legitimate and unimportant. Highly legitimate product questions are distributed among classes 2 and 3. Trying to identify the logic behind this distribution, classification parameters have been confirmed: class 2 questions are asking for product attributes that might be less relevant to customers in the product choice process (such as the question asking for the type of hood on the jacket or the carrier cord of the camera) while questions in class 3 address product attributes with more choice relevance (such as color and material of the jacket or weight and zoom of the camera).

Classif1 * cat2 Crosstabulation

			cat2				Total
			Pd	Peip	Pepr	U	
Classif1	2,00	Count	14		3	1	18
		% within cat2	33,3%		13,0%	7,7%	16,1%
		% of Total	12,5%		2,7%	,9%	16,1%
	3,00	Count	26		13	2	41
		% within cat2	61,9%		56,5%	15,4%	36,6%
		% of Total	23,2%		11,6%	1,8%	36,6%
	4,00	Count		1		1	2
		% within cat2		2,9%		7,7%	1,8%
		% of Total		,9%		,9%	1,8%
	5,00	Count	2	5	7	6	20
		% within cat2	4,8%	14,7%	30,4%	46,2%	17,9%
		% of Total	1,8%	4,5%	6,3%	5,4%	17,9%
	7,00	Count		28		3	31
		% within cat2		82,4%		23,1%	27,7%
		% of Total		25,0%		2,7%	27,7%
Total		Count	42	34	23	13	112
		% within cat2	100,0%	100,0%	100,0%	100,0%	100,0%
		% of Total	37,5%	30,4%	20,5%	11,6%	100,0%

Table 3: Questions Type and Leg x Imp Classes

Looking into the perception of person oriented questions it is not surprising to note that people attribute little legitimacy and importance to those questions that only focus on the individual and obviously do not contribute to product or service delivery (*peip*). Asking for age, address, hobbies or other information therefore does not seem appropriate in an online context if there is no reason for it. On the other hand, there is a relatively high acceptance (56,5%) of questions that even though focusing on the person do have a connection with product selection (*pepr*-questions). This implies that customers in many cases do not feel annoyed if they are asked personal questions as long as these relate to the product context. In fact, none of the *pepr*-questions have been perceived as totally illegitimate or unimportant. Looking more closely into those *pepr*-questions that are perceived as highly legitimate it seems that asking people what they 'prefer' is perceived more legitimate and important (class 3) than asking them 'how important' they perceive one or the other product feature to be (class 5). This finding could be an interesting area of future research. The data material in the present study is not large enough to sufficiently investigate this issue.

Finally, questions concerning usage (u) need some recognition: those that relate somehow to features of the product (like motives you want to capture with the camera) are perceived as sufficiently important and legitimate (class 5). On the other hand, those that lack a link to product selection are perceived as rather illegitimate and unimportant. Figure 4 demonstrates some of the relationships found.

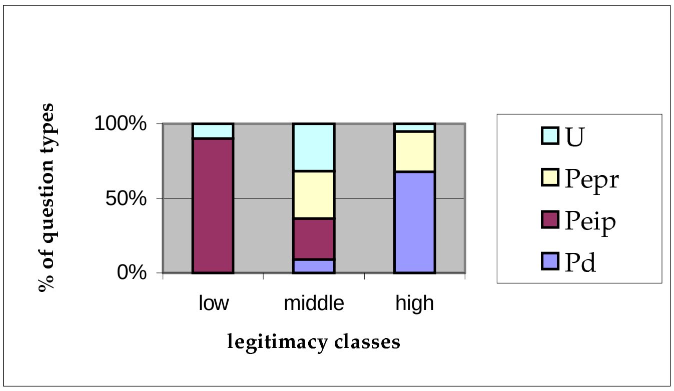

Figure 4: Relationship between Legitimacy and Question Type

4.3. Final Definition of Overall Model

Formal co-linearity diagnostics as well as the strong linear relationship between *Leg* and *Imp* depicted in Figure 3 led us to the conclusion that the validity of results obtained for the original fixed effects model (1) might be questionable. We therefore re-specified the model estimating a simultaneous equation model where in addition to the direct effects of *Leg*, *Imp* and *Diff* on *PCIC* we included a linear relationship between *Leg* and *Imp*:

$$\begin{aligned} PCIC_{ij} &= \beta_0^{IC} + \beta_1^{IC} Leg_{ij} + \beta_2^{IC} Imp_{ij} + \beta_3^{IC} Diff_{ij} + \varepsilon_{ij}^{IC}, \\ Leg_{ij} &= \beta_0^{Leg} + \beta_1^{Leg} Imp_{ij} + \varepsilon_{ij}^{Leg}. \end{aligned} \tag{2}$$

Again dummy variables were used to control for individual differences. As was shown, significant differences exist between product related questions (group A: *pd*, *pepr*) and more or less unrelated questions (group B: *u*, *peip*) as far as the perception of legitimacy and importance is concerned. Based on (2) we therefore estimated two group-specific models in addition to one representing the total sample. Maximum Likelihood estimates for the model parameters (Table 4) have been generated by M*plus* [19], a software for the estimation of mean- and covariance structure models (widely known as SEM). Because of the small number of respondents one might be tempted to reject the application of this methodology in our study. To put this objection into perspective the following facts should be taken into consideration. First, although sample size is 38 the number of observations is much higher since we collected multiple data (112 questions) for each respondent. This results in a total sample size of 4,256 observations. Second, our analysis does not correspond to typical SEM applications where latent variables with multiple indicators are involved. It therefore is questionable if general minimum sample size recommendations (100 - 200) or rules of thumb developed for these more complex models apply also to our study. Third, the ratio of sample size (4,256) to number of free parameters (82) is 52:1, which is considerably higher than the ratio of 10:1 suggested by Bentler and Chou [6] to obtain valid parameter estimates and standard errors.

Since model (2) has one degree of freedom in addition to the multiple correlation coefficient $\hat{R}^2$ alternative overall fit measures for covariance structure analysis have been used (for the interpretation of these fit statistics see for example [16]). As can be seen from Table 4, results for the total sample as well as for group A show an excellent fit according to the RMSEA [8,15]. However, we should bear in mind that because of the extremely low degrees of freedom fit statistics have low power [17]. This might explain the wide confidence intervals for RMSEA. In contrast, results for group B definitely represent a borderline case as indicated by a fairly high RMSEA of .070. Therefore the estimates for this group should be interpreted with particular caution.

Coefficients of the total sample clearly show that the effect of *Imp* on PCIC has been underestimated by the original single-equation fixed effects model (1). Although the direct effect is still insignificant, the total effect (-.499) is only moderately smaller than the legitimacy effect (-.559). The impact of perceived importance on information costs is thus obviously predominantly mediated by its influence on perceived legitimacy.

Since the two group-specific models display some significant differences they will be interpreted in more detail: Just as for the total sample the most important driver of PCIC in both groups is the perceived legitimacy of an information request. *Imp* drives PCIC predominantly via its influence on *Leg*. Only for person-related questions (group B) a small direct effect seems to be present. Compared to the direct effect of *Leg* and the total effect of *Imp,* the difficulty to answer a question is obviously perceived as less costly by respondents. As might have been expected from the preceding analysis of the *Leg-Imp* relationship (Figure 3), *Imp* has a much stronger influence on *Leg* in group B than in group A. Likewise the effect of *Leg* on PCIC is stronger in group B. As far as *Diff* is concerned, there are only minor differences between the two groups.

Overall model fit		
Total sample	Group A	Group B
$\chi^2_{(1)} = 1.86$ RMSEA = .014 RMSEA 90% CI (.000, .046) $\hat{R}^2_{IC} = .622$ $\hat{R}^2_{Leg} = .739$	$\chi^2_{(1)} = 4.34$ RMSEA = .037 RMSEA 90% CI (.007, .075) $\hat{R}^2_{IC} = .481$ $\hat{R}^2_{Leg} = .594$	$\chi^2_{(1)} = 9.74$ RMSEA = .070 RMSEA 90% CI (.035, .113) $\hat{R}^2_{IC} = .693$ $\hat{R}^2_{Leg} = .735$

Parameter Estimates				
		Total Sample	Group A	Group B
Explanatory variables	**Parameter**	**Dependent variable: *PCIC***		
Intercept	β_0^{IC}	6.250	4.569	6.274
Leg	β_1^{IC}	-.559 (.017) ***	-.397 (.022) ***	-.457 (.027) ***
Imp	β_2^{IC}	Direct effect		
		-.010 (.017)	.003 (.019)	-.055 (.029) *
		Total effect		
		-.499	-.232	-.437
Diff	β_3^{IC}	.138 (.014) ***	.182 (.016) ***	.159 (.020) ***
		Dependent variable: *Leg*		
Intercept	β_0^{LEG}	1.289	3.737	.714
Imp	β_1^{LEG}	.875 (.009) ***	.591 (.013) ***	.839 (.015) ***

( ) standard error; ***p < .01; *p < .10

Since the data consists of partially dependent observations, controlling for these dependencies might lead to slightly lower levels of significance.

Table 4: Model Results for Simultaneous Equation Models with Fixed Effects

5. Impact of Model Results

Summing up, measures to manipulate PCIC through strategic interface design should foremost concentrate on higher levels of legitimacy and importance of information requests as these variables have a higher impact on PCIC. On this background, empirical findings allow for a critical discussion of current EC communication practices and at the same time lead to some suggestions of improvement.

Today, most EC websites are only asking users for desired product attributes (pd) (e.g. product configuration engines on manufacturers sites or product search engines on infomediary sites) or they ask them to fill out lengthy online questionnaires which mostly contain personal questions (peip). Very few sites start to include questions on usage (u) and nobody is communicating with users yet on general product expectations (pepr).[4] As was shown above, however, users do accept personal questions as long as they relate to the product context (pepr-questions). For example, asking a consumer whether he prefers trend models when choosing a jacket is initially a personal question, because it contains information on the consumer's general attitude towards fashion. As such it has considerable value for sellers, because they directly learn about their buyer's preference. However, the information unit also serves directly to recommend the right type of product to the client by respecting the degree of trendiness of different models in the electronic choice process. Strictly speaking, most marketers realize opportunity cost of information today if they do not take advantage of the potential knowledge accumulation they can realize with pepr-questions.

Additionally, as can be seen from graph A in Figure 3, *pepr*- as well as *pdd*-questions are less driven by the *Imp* factor than personal- or usage oriented questions (graph B has a steeper slope than graph A). This finding implies that as questions become slightly less important for the customer, their legitimacy is not decreased to the same extend. Taking advantage of this relationship means that marketers could ask customers *pdd*- or *pepr*-questions that even though less relevant to the buyer are still important for product enhancement purposes. For example, asking consumers what type of closing mechanism they prefer for compact cameras might not be too relevant a question for most buyers. Yet, for manufacturers of compact cameras this information is highly valuable for product design decisions.

Considering in contrast the impact of *Imp* on the perceived *Leg* of *peip*- and *u*-questions it becomes obvious that marketers have to be careful to employ this type of question in web sites. However, especially *u*-questions have the potential to be accepted if their importance for the choice process justify them.

6. Conclusion and Outlook

The contribution of this article is that is raises awareness for e-privacy, or more precisely, for PCIC as a potentially important search cost dimension in electronic markets. Based on empirical data, a functional model is presented which shows that PCIC can be explained to some extend by the three factors of perceived legitimacy, importance and difficulty of an information request. This gives marketers an orientation in how to design online communication more 'consciously' with regard to PCIC. Relating different types of interface questions (*pdd*, *pepr*, *peip* and *u*) to the main drivers of PCIC has revealed the opportunity for marketers to ask more person oriented questions in online purchase contexts than is currently the case. Finally, the model presented might be a starting point to compare the PCIC perception of different communication catalogues. Doing so, strategic interface design can follow suit. One option is to decrease PCIC. In this case, interface design should foremost concentrate on higher levels of legitimacy and importance of information requests. Since product-related questions by their nature already score high on legitimacy and importance, improvements on these dimensions are much harder to realize for them than for person-related questions. Or, in contrast, PCIC is consciously maintained at a higher level. Awareness for higher PCIC could then, however, be the basis for the definition of appropriate returns.

We are aware of the limitations of the current research. Especially the small number of subjects restricts a broad generalization of the results presented in this paper. Also the overall model fit for group B suggests that besides the three factors identified other factors play a role in the evaluation of PCIC. Still we feel that with this work we are presenting an innovative approach to evaluate private information provision on the Internet and also help to raise awareness for this factor. Moreover, a number of open questions also become obvious for future research programs: For example, if marketers wanted to offer appropriate returns to consumers it is vital for them to know how consumers actually evaluate those (e.g. web miles, free services, cash etc.). What is the exchange value of private information? Also, what are the personal factors potentially driving this exchange value? The influence of the personality and personal experiences are a factor only marginally recognized in the model presented above through the employment of dummy variables. No insights have been gained on how personal traits such as product knowledge, Internet experience or privacy actually play on the perception of PCIC. Also, it cannot be excluded that the order in which questions are asked on a web site influences the perception of PCIC. Most importantly it is questionable whether consumers even though perceiving a certain cost level do act accordingly, thus answering questions only up to a cost level *x*. In fact, other variables such as trust in the online vendor, the uncertainty related to product choice, the perception of the search engine etc. are all variables that might lead a user to answer more questions than would be intuitively suggested by PCIC. Finally, PCIC should be investigated in relation to other search cost variables. For example, it would be interesting to investigate the relationship between time cost and PCIC and how they interrelate with each other in the formation of overall online search cost.

[4] For a critical review of currently employed interface agents in EC websites see also: [25]

[3] Even though there is a lot of progress in recommendation engines that do not require any active input from the user there will always be product categories for which considerable exchange between buyers and sellers is necessary (e.g. trust goods). Already today, high-quality recommendations made by infomediaries are a service paid for mostly by customer information (e.g. Active Buyers' Guide). Also long term consumer agent projects, such as the REA project at MIT are envisioning dialogues between buyers and sellers very similar to the real world. Here, even more information, especially personal information, will be revealed by consumers.

[4] Traditional theories of information value have a different perspective on value creation: While they are concerned mostly with the benefits for the recipient of information compared to the production cost of this benefit, we are more interested in the cost of the provision of an additional unit of information while at the same time this provision leads to no measurable production cost.

References

[1] Ackerman, M.S., L.F. Cranor and J. Reagle, "Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences", in: *Proceedings of the ACM on E-Commerce*, Denver, Colorado (1999)

[2] Badenoch, D. et al., "The value of information", in: The Value and Impact of Information, eds. by Feeney, M. and Grieves, M., British Library Research, East Grinstead, (1994)

[3] Bäumler, H., E-Privacy – Datenschutz im Internet, Braunschweig/Wiesbaden, (2000)

[4] Beatty, S.E. and S.M. Smith, "External Search Efforts: An Investigation Across Several Product Categories", Journal of Consumer Research, vol.14, pp. 83-95, (1987)

[5] Belsley, D.A., E. Kuh and R.E. Welsch, "Regression Diagnostics: Identifying Influential Data and Sources of Collinearity", New York: Wiley, (1980)

[6] Bentler, P.M. and C.P. Chou, "Practical Issues in Structural Modeling", in: Sociological Methods and Research, vol. 16, pp. 78-117, (1987)

[7] Bettman, J.R., "An Information Processing Theory of Consumer Choice", Reading, MA: Addison-Wiley, (1979)

[8] Browne, M.W., Cudeck, R., "Alternative Ways of Assessing Model Fit", in: Testing Structural Equation Models eds. by Bollen, K and Long, J., Newbury Park: Sage, 136-162, (1993)

[9] Chang, A., Kannan, P.K. and Whinston, A.B., "The Economics of Freebies in Exchange for Consumer Information on the Internet: An Exploratory Study", in: International Journal of Electronic Commerce, Vol.4, No.1, pp. 85-102, (1999)

[10] Dyson, E., "Privacy Protection: Time to Think and Act Locally and Globally", Release 1.0, April, (1998)

[11] Evans, P. and Wurster, T.S., Blown to Bits – How the New Economics of Information Transforms Strategy, Boston, MA, (2000)

[12] Feeney, M. and Grieves, M., The Value and Impact of Information, British Library Research, East Grinstead, (1994)

[13] Hagel III, J. and Rayport, J.F., „The Coming Battle for Customer Information", in: Harvard Business Review, January-Febuary, (1997)

[14] Hine, C. and J. Eve, "Privacy in the Marketplace", in: The Information Society, vol. 14, pp. 253-262, (1998)

[15] Hu, L. and P.M. Bentler, "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives", in: Structural Equation Modeling, vol. 6, pp. 1-55, (1999)

[16] Jöreskog, K.G., "Testing Structural Equation Models", in: Testing Structural Equation Models, eds. by Bollen, K, Long, J., Newbury Park: Sage, 294-316, (1993)

[17] MacCallum, R.C., Browne M.W. and H.M. Sugawara, "Power Analysis and Determination of Sample Size for Covariance Structure Modeling", in: Psychological Methods, vol. 1, pp. 130-149, (1996)

[18] Moorthy, B. Ratchford and D. Talukdar, "Consumer Information Search Revisited", in: Journal of Consumer Research, vol.23, no. 4, pp. 263-277, (1997)

[19] Muthén, L.K., Muthén, B.O., Mplus User's Guide, Los Angeles: Muthén & Muthén, (1998)

[20] Neufeldt, V. and Guralnik, D.B., Webster's New World Dictionary, 3rd College Edition, New York, (1998)

[21] Pew Internet & American Life Project, Trust and Privacy Online: Why Americans Want to Rewrite the Rules, 2000-8-20, http://pewinternet.org/reports/toc.asp?Report=19

[22] Punj, G. and R. Staelin, "A Model of Consumer Information Search Behavior for New Automobiles", in: Journal of Consumer Research, vol. 9, pp. 366-380, (1983)

[23] Shapiro, C. and Varian, H.R., Information Rules – A Strategic Guide to the Network Economy, Boston, MA, (1999)

[24] *Spiekermann, S., Grossklags, J., Berendt B.*, "E-privacy in 2nd generation E-Commerce: privacy preferences versus actual behavior"; submitted to ACM conference on EC; October 2001; can be downloaded from: http://www.wiwi.hu-berlin.de/~sspiek/phdresearch.html

[25] Spiekermann, S. and Corina P., “Motivating Human-Agent Interaction : Transferring Insights from Behavioral Marketing to Agent Design”, in: *Proceedings of the 3rd International Conference on Telecommunications and Electronic Commerce, ICTEC3*, 2000, pp.387-402
[26] Srinivasan, N. and B.R. Ratchford, “An Empirical Test of a Model of External Search for Automobilies”, in: Journal of Consumer Research, vol. 18, pp. 233-242, (1991)
[27] Stigler, G., “The Economics of Information”, in: Journal of Political Economy, vol. 69, no.3, pp. 213-225, (1961)
[28] Volokh, E., Personalization and Privacy, in: Communications of the ACM, Vol.42, No.8, pp.84-88, (2000)
[29] Westin, A., “Harris-Equifax Consumer Privacy Survey”, Atlanta, GA:Equifax Inc. (1996)