Development of a reactance scale for human–computer interaction

Patrick Ehrenbrink¹ · Sebastian Möller¹

Received: 29 March 2017 / Published online: 1 February 2018
© The Author(s) 2018. This article is an open access publication

Abstract

Psychological reactance is an effect known from social psychology. Its relevance has been shown in marketing and health communication but has not yet been fully introduced to the field of human–computer interaction. This work describes the development and validation of the reactance scale for human–computer interaction (RSHCI). The RSHCI is designed to measure state reactance in the context of HCI research and is the first questionnaire to address the topic in this context. First, an online study was conducted to find the most reliable items by means of principal component analysis and structural equation modeling. Afterwards, a laboratory experiment was conducted to test the validity of the questionnaire that was obtained in the online study. The result is a highly reliable questionnaire that resembles the structure of psychological reactance as it was identified in related research and also contains a confirmatory factor.

Keywords Psychological reactance · HCI · Intelligent personal assistant · Persuasive systems

Introduction

In recent years, many products that show highly complex behavior and functionality appeared in consumer electronics.

Among this, one can find many persuasive systems that are intended to persuade their users into e.g. doing more sports or eat healthier. Those systems can use a variety of strategies to persuade their users and thereby apply different levels of persuasive pressure or motivation. If the pressure is too high, however, the user could perceive the persuasive attempt as a threat to his or her personal freedom. Causing freedom threats should be avoided by any system or service because they can cause psychological reactance. Reactance is an affective and motivational state of the user. It might result in an unwanted decline of acceptability of a product [1] or less compliance [2]. This work describes the development and introduces a questionnaire that can measure levels of a user's psychological reactance. An efficient method that is able to measure reactance can support developers and researchers to identify devices, applications, services or interaction strategies that are prone to induce reactance

and act against it before having to cope with unnecessarily low user acceptance.

Psychological reactance

Personal freedom is one of the inherent needs of people. If their freedom is threatened, they can react exceedingly resolute and sometimes inappropriate, in an attempt to restore or preserve that freedom. Such behavior is then called reactance or reactant behavior. The concept of reactance was first introduced by Brehm [3]. Reactance can originate from a perceived freedom threat. Such a freedom threat can be any event or occurrence that prevents or inhibits people from exercising their freedom. This might be some form of prohibition, but also persuasive attempts by others. Reactant behavior is the attempt to reestablish or preserve personal freedom against the perceived threat. Even scarceness of products can cause reactant behavior and make people wanting to buy products that they would not be interested in, without them being scarce. In that case, the scarceness creates the anticipation of not being able to buy that product anymore. This threat to the freedom of “not being able to buy the product anymore” is then prevented by buying that product while it is still available. The importance of reactance has been recognized in areas like marketing [2] or medicine [4]. Especially the effect of scarceness is used

✉ Patrick Ehrenbrink
patrick.ehrenbrink@tu-berlin.de

¹ Electrical Engineering and Computer Science, Quality and Usability Lab, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany

widely in marketing and sales, e.g. in advertising slogans like “for a limited time only” or “exclusive”.

An example from the field of human–computer interaction could be a persuasive cooking assistant. A cooking assistant that monitors the eating behavior of its users and then provides tips on how to eat healthier. A user of such a system could use it to search for recipes for her or his dinner. If the cooking assistant proactively claims something like “You should eat a healthy salad, you had meat every day this week.”, the user might perceive this as a threat to his or her freedom of choice. According to reactance theory [3], the user might engage in reactant behavior and try to restore the perceived loss of freedom of choice by deliberately choosing to eat meat, instead of a salad. Even if she or he might have intended to have a salad in the first place.

Reactance can be addressed by two different viewpoints. The first one is trait reactance. Trait reactance means reactance as a personality trait. Trait reactance of a person is stable over time and typically invariant to a specific situation. It can be regarded as a part of the person’s personality and can be assessed by a number of different questionnaires, such as Hong’s Psychological Reactance Scale [5, 6] or the Therapeutic Reactance Scale [7]. State reactance, on the other hand, is a time and situation-dependent affective and motivational state, where reactance behavior is triggered by a stimulus. Unlike trait reactance, state reactance can be moderated by applying strategies. This will be explained further in Sect. 2.1.

Related work

This section will shortly present the state of the art in reactance research in general and will then discuss related work that utilizes reactance in the context of HCI in more detail. As explained in the previous section, there are two different concepts of reactance. Reactance as a personality trait is relatively stable over long periods of time or even a whole life, whereas reactance as an affective and motivational state varies over time and is situation-dependent.

Evoking and avoiding psychological reactance

If one wants to develop a questionnaire that can measure a certain construct, it needs to be constructed and tested with a dataset that includes different levels of the construct which is supposed to be measured. For the current work, this means that stimuli are needed which evoke different levels of reactance, e.g. high and low. A common way of evoking psychological reactance under laboratory conditions is by presenting participants with two different stimulus texts. One stimulus text is supposed to pose a high threat to the perceived freedom of the participant. This is usually evoked

by the use of a commanding tone and overly persuasive formulation of the text, e.g. “Flossing your teeth is absolutely mandatory for keeping healthy teeth. You have to do this at least three times a day!”. On the other hand, a low freedom threat stimulus usually emphasizes the individual’s freedom of choice and avoids commanding tone: “Flossing can significantly reduce caries. You can aid the health of your teeth by flossing. It is your choice.” According to reactance theory [3, 8], a freedom threat can evoke reactance. Therefore, reactance levels are expected to be higher in the high threat condition, compared to the low threat condition. Similar stimuli have been used by Dillard and Shen [2] and others [9, 10]. Furthermore, a meta-analytic review of several studies which also used high threat and low threat texts as stimuli to manipulate reactance was conducted by Rains [11]. It can be argued, that high freedom threat and low freedom threat texts provide widely accepted stimuli to manipulate reactance.

Measuring psychological reactance

The first questionnaire (18 items) that was developed to measure trait reactance was the Fragebogen zur Messung der Psychologischen Reaktanz (FMPR) introduced by Merz [12]. That questionnaire was later translated into English, modified and published as Hong’s Psychological Reactance Scale (14 items) (HPRS) by Hong and Page [5]. Hong and Fedda [13] published a revised version of HPRS (11 items). Hong and Fedda describe the HPRS as composed of four factors, namely “Emotional Response Toward Restricted Choice”, “Reactance to Compliance”, “Resisting Influence from others” and “Reactance Towards Advice and Recommendations”. An example item from the revised HPRS is “Regulations trigger a sense of resistance in me”. The multi-factor structure of the HPRS is still a matter of debate. It is argued that it should better be treated as a single factor measure [14, 15]. Another questionnaire is the Therapeutic Reactance Scale (TRS) (28 items) which was published by Dowd et al. [7]. It addresses reactance with the two factors “Behavioral Reactance” and “Verbal Reactance”. An example item of this scale is “If I receive a lukewarm dish at a restaurant, I make an attempt to let that be known”. According to the number of citations, the refined version of HPRS is the most widely used or discussed scale for assessing trait reactance.¹

In recent years, the development of measurement tools has focused on state reactance. Several measurement techniques have been introduced in the literature.

Probably the most prominent method was introduced by Dillard and Shen [2]. Dillard and Shen compared several

¹ Citations according to Google Scholar December 20, 2017: HPRS (old version): 190 citations; HPRS (refined version): 350 citations; FMPR: 92 citations; TRS: 296 citations.

models of reactance and concluded that the “intertwined model” is the one that fits the observed data best. The intertwined model describes reactance as consisting of two components, namely anger and negative cognitions. They claim that anger and negative cognitions are entangled in a way that the individual contribution to the user’s state reactance, or better, to the user’s reaction to the freedom threat cannot be known. They assess state reactance by measuring those two components separately, using two different techniques. Anger is measured with a four-item questionnaire which used self-report items. It asks for the extent that people feel irritated, angry, annoyed and aggravated. Negative cognitions, on the other hand, are collected and then rated via a thought-listing task [2]. Users are asked to write down whatever comes to their minds. Afterwards, the texts are partitioned into separate thoughts. The thoughts are then rated as either positive, neutral or negative by the participants. Only the negative thoughts are counted. A major drawback of this technique is that the amount of collected thoughts is strongly dependent on the participant and that the relation of positive to negative thoughts is not assessed. If person A would come up with 50 thoughts, two of which are negative and 48 are positive, while Person B comes up with one negative thought only, person A’s measure of negative cognitions would still be regarded as more negative. Also, there is the problem of social desirability. Usually, though listing is done under close supervision with the experimenter. As the participants have to come up with the thoughts by themselves, their responses will probably be biased towards what they assume the experimenter wishes as an outcome [16].

Another measurement technique is a four-item self-report scale that was developed by Lindsey et al. [17] and supposedly uses modified items from Hong’s Psychological Reactance Scale [5, 6]. The validity of the scale is questionable because, at close inspection, the used items do not resemble those of Hong’s Psychological Reactance Scale very much. Probably the most salient similarity is the use of the phrase “It irritates me...” in one item. Two of the other items can be regarded as highly redundant, namely “I do not like that I am being told how to feel about bone marrow donation” and “I dislike that I am being told how to feel about bone marrow donation” [17].

The third method is the Reactance Restoration Scale, a 12 item questionnaire developed by Quick [18]. It consists of three cloze texts, that have to be answered by using a four-item semantic differential. The Reactance Restoration Scale aims at measuring boomerang effects that occur when people get reactant, rather than reactance or supposed components of reactance. The items have to be tailored to a specific situation, e.g. “Right now, I am ____ to (exercise/use sunscreen the next time I am exposed to direct sunlight for an extended period of time [greater than 15 min])” [18].

A fourth method is the Salzburger State Reactance Scale [19]. It is a Likert-type, ten-item questionnaire that shows a three-factor structure. Two of the factors closely resemble the supposedly underlying constructs of reactance, anger, and negative cognitions, while the third factor aims at aggressive intentions towards the stimulus. This factor could be interpreted as a confirmatory factor. This scale has a rather focused design towards a specific context since the items would have to be modified quite strongly to fit another context. This is especially true for the items of the factor that resembles negative cognitions. One example of such an item is “Do you think that the landlord also shows discriminatory behavior in other areas?” [19].

Psychological reactance in HCI

Reactance was only scarcely used as a dependent variable in HCI research, despite the fact that its importance seems to be acknowledged by researchers. For example, it was considered in the design phase of a product which was then designed in a way that it would induce less reactance. This was done by implementing features that were described as naive, understanding or ironic [20]. In another case, reactance was used as a possible explanation for observed results, even though it was not measured [21].

An attempt to avoid reactance was suggested by Hasenzahl and Laschke. They argue that technology that can persuade users necessarily causes friction, which then could result in reactance. As a solution, they propose to design the friction in a way that it is more acceptable for users. They provide an example of a persuasive ice-cream bowl that is denying its users to watch TV while eating ice cream. Since the users are still allowed to eat ice-cream, the restriction of not watching TV should become more acceptable [22].

Apart from this, some studies exist that measured reactance either indirectly or directly. Those will be introduced shortly in the following.

A study in which reactance was measured indirectly was performed by Lee and Lee [23], who inferred on reactance by measuring the strength of a perceived threat and then concluded that a strong threat must result in higher reactance. This is in line with the wide use of high freedom threats as triggers for reactance in literature. A similar approach was pursued by Murray and Häubl, who varied freedom of choice and then inferred on reactance from this [24]. For many situations, inferring on reactance by the presence of known triggers is a legitimate procedure. However, results might not be as reliable as with a direct measurement. A direct measure of reactance would provide more certitude in this matter.

Two studies could be identified that measured reactance directly in an HCI context. Kwon and Chung used the refined version of Hong’s Psychological Reactance Scale, hence

measuring trait reactance, in a study about online shopping recommendations [25]. Roubroeks et al. used the Reactance Restoration Scale [18] to assess state reactance [26] and found that the amount of reactance was higher if a persuasive stimulus was accompanied by an image or video of a robotic agent, compared to text only.

The relatively low presence of the concept of reactance in HCI research could be due to the lack of devices or services that are prone to trigger high reactance in the past. However, in recent years human-like aspects and persuasive features have found their way into everyday services and applications. Examples for this are intelligent personal assistants or user-adaptive shopping services. The recent rise of such services and applications justifies a stronger focus on concepts like reactance in the HCI community. In order to enable a wide use of the reactance-concept, its measurement and interpretation should be straightforward. This means that the measurement technique should not require high expertise (thought listing) or considerable adaptation of questionnaire items. Therefore, the goal of this work is to provide a measurement tool for state reactance that is straightforward to use, short and easy to analyze. The steps that were taken to reach this goal are presented in the following.

Development

Developing a questionnaire requires a large set of data that is usually collected in many studies or above different conditions within the same study. For this work, an online experiment was created in which all participants were assigned to one of two conditions (between subjects design). Each participant was exposed to one of two stimulus texts, describing a smart home, that was meant to either trigger reactance or not. The texts can be found in Tables 4 and 5. To create enough variance in the data, it was essential to trigger reactance within one half of the test population, while avoiding it in the other half. The two stimulus texts have been created to accomplish this. The first one is supposed to pose a high threat to freedom. It describes situations in which responsibilities are assigned to the participant and is formulated in a commanding tone. Further, some of the described components were selected because they describe an authoritarian element that would further act as a threat to personal freedom, such as a ‘moral adviser’. The other text is supposed to pose a low threat to freedom. It was designed so that it emphasized the participants’ freedom of choice through its employment of the “but you are free to accept or to refuse” technique. An example of this is ‘you can choose if you want to use it or not’. This method was described by Guégen and Pascual [27] and demonstrated higher compliance if people were asked to participate in a study. As compliance increases, it can be assumed that reactance decreases since non-compliance is a reactant behavior.

Item generation

An initial set of items was created for the current study. The items were generated in a three-step process. In a first step, phrases for the items were collected from three different sources. The first source was the existing questionnaires that measure reactance. In particular, we used the ‘Anger’ scale [2], ‘Hong’s Psychological Reactance Scale’ [5], ‘Fragebogen zur Messung der psychologischen Reaktanz’ [12]² and paraphrased items that could potentially be formulated to address a specific trigger. This resulted in a total number of 15 unique phrases. The second source was anonymous user comments from websites, particularly ‘<http://www.zeit.de>’ and ‘<http://www.heise.de>’ among a number of others. Anonymous user comments of websites were chosen to avoid formulation biases by social desirability or social anxiety effects that would likely appear in interviews or focus groups. It has been shown that people tend to show less social desirability and social anxiety when they are communicating via the internet and remain anonymous [28]. Formulations from the comments were selected when they supposedly reflected either one of the components of psychological state reactance as proposed by Dillard and Shen [2], namely anger and negative cognitions, or when they indicated that the person felt his or her personal freedom threatened. The formulations were collected by two researchers independently and resulted in a total number of 15 unique phrases. A third source was the outcome of a brainstorming session in which six HCI experts that were knowledgeable about the concept of psychological reactance participated. The brainstorming session resulted in a total number of 22 unique phrases. Following the suggestions from some of the HCI experts, also positively formulated phrases were included.

In a second step, two researchers consolidated the 52 collected phrases. This step included the removal of double (or triple) phrases from the three different sources. Also, offensive phrases were removed. After consolidation, a set of 32 phrases remained.

In the last step, 37 items were generated by two researchers as sentences that used the 32 phrases in a way that they would address the system under evaluation. All 37 items and the concepts that they address can be viewed in Table 6.

Procedure

All participants had to register themselves with their email address prior to the experiment. This was done to reduce the risk of participants taking part in the experiment more often than once. After registration, participants would be

² ‘Hong’s Psychological Reactance Scale’ was developed based on the ‘Fragebogen zur Messung der psychologischen Reaktanz’.

provided with a link to the survey via email. After starting the survey, participants had to state their age. They were then assigned to either the high-threat condition or the low-threat condition based on their age. Participants who stated an even number as their age were assigned to the high-threat condition, participants with an odd age were assigned to the low-threat condition. Afterwards, each participant was presented with either the high-threat stimulus text or the low-threat stimulus text and was instructed to read it thoroughly. After reading, participants could enter the next page by clicking on a button. They were then presented with the 37 initial items of the questionnaire. Each item had to be answered according to a 5-pt Likert-Scale reaching from “Strongly disagree” to “Strongly agree”. The respective stimulus text was still visible on this page. Only after answering all 37 items, participants would be able to enter the next page. At the end of the survey, participants were asked to answer five control questions to ensure that they had actually read and understood the stimulus text. The questions included two free text questions and two multiple choice questions about the content of the text and one question asking directly if the participant had read the text thoroughly. Afterwards, participants could choose if they wanted to take part in a lottery in which they could win one of three prizes of 50 €, each.

Participants

Participants were recruited via a participant database of Technische Universität Berlin. The database includes about 3000 persons. Three times 50 € were drawn under all participants who took part in the experiment. In total, 448 participants took part, 395 of which completed all 37 items for measuring reactance. The mean age of those participants was 29.32 years (std. 9.66 years). Further details, such as gender or occupation were not assessed.

Analysis

Prior to the actual analysis, the responses of all participants were reviewed to filter out invalid sets. Sets were regarded as invalid if either not all questions were answered (participants aborted the survey), if the responses were obviously non-sense (e.g. answered ‘Strongly Agree’ for each question) or when participants did not pass the control questions. A factorial analysis was performed on the 342 remaining responses. In the first step, the dataset was split into two subsets at random. This resulted in one dataset with 162 subjects (dataset_A) and one dataset with 179 subjects (dataset_B). Afterwards, a factor analysis was performed according to the method applied by Wechsung [29]. First, a maximum likelihood factor analysis was performed on dataset_A to select the most appropriate items for each factor. Afterwards, a confirmatory factor analysis was performed on dataset_B.

Table 1 Indices of the intermediate factors for Dataset_A after the exploratory phase of the factor analysis

Factor	α	% of variance	Goodness-of-fit test		
			χ^2	df	Sig.
‘Antipathie’	0.898	71.486	6.507	5	0.260
‘Wut’	0.945	75.297	15.539	14	0.342
‘Autonomie’	0.878	67.714	8.978	5	0.110

Factors ‘Antipathie’ (Negative Cognitions) and ‘Wut’ (Anger) are in line with the intertwined model. ‘Autonomie’ is the confirmatory factor for a perceived threat-to-freedom

Factor construction

Item difficulty was calculated for all items according to Moosburger and Kelava, who proposed to keep items between 20 and 80 [30]. Since item difficulty varied between 40.73 and 77.17 for all items, no items were removed in this step.

In the next step, all items were assigned to one of the three factors (‘Anger’, ‘Negative Cognitions’ and ‘Threat’) that they were intended to represent. Item discriminant indices were used to further reduce the item set. This was done using the ‘Corrected Item-Total-Correlation’ [31] measure in IBM SPSS Statistics 23 [32]. Only items that showed a value above 0.6 were kept for further analysis.

Next, a factor analysis with maximum likelihood as extraction method and varimax rotation for exactly one factor was performed on all items that were supposed to represent one factor. When the goodness-of-fit test showed a significant result, the procedure was performed again for a fixed number of two factors. Then, only items that belonged to the one factor with more than 50% of explained variance were kept. This procedure was proposed by Homburg and Giering [33]. After this step, there were five items remaining for the factor ‘Negative Cognitions’, seven items remaining for the factor ‘Anger’ and six items remaining for the factor ‘Threat’. This first phase resulted in a three-factor questionnaire, consisting of 18 items. In consideration of the remaining items being in German, the factor ‘Negative Cognitions’ was renamed to ‘Antipathie’ (GER), the factor ‘Anger’ was renamed to ‘Wut’ and the confirmatory factor to ‘Autonomie’. Cronbach’s alpha, explained variance and fit-indices calculated on Dataset_A can be reviewed in Table 1.

Confirmatory factor analysis

In the second phase, a confirmatory factor analysis was performed on Dataset_B to validate the factors of the new reactance scale. This was done by structural equation modeling, using IBM SPSS Amos 24.0.0 [34]. Structural equation modeling is a method where a model based on theoretical assumptions [35], in this case, the formation of specific factors by a number of items, can be built and then tested

on empirical data. The sample size of the test data was 179. For a structural equation model, sample size should be $N \geq 100$ minimum [36]. The model can be seen in Fig. 1. As indicated by the two-sided arrows, all three factors were allowed to correlate. This is because the factors ‘Wut’ and ‘Antipathie’ are both assumed to be sensitive to a perceived freedom threat. As ‘Autonomie’ is representing that freedom threat, it is also correlated to the two other factors. After implementation, Amos was used for calculating a number of fit indices that were used to judge if the model was appropriate and represented the factor structure of the items also on Dataset_B. The fit indices and the respective thresholds for meeting the requirements were selected in accordance with the criteria suggested by Homburg and Giering [33]. The criteria include the following fit indices:

- Comparative Fit Index (CFI) is a comparative method in which the model fit of a model is compared to a hypothesized model. CFI ranges from 0 to 1 and a higher number means a better model fit [37]. The criterion applied here was $CFI \geq 0.95$.
- Root mean square error of approximation (RMSEA) takes the error of approximation into account. It is a measure of how bad a model fits the data. Hence, a smaller RMSEA means a better model fit. The values range from 0 to 1 [38]. The criterion applied here was $RMSEA \leq 0.08$.
- Adjusted Goodness of Fit Index (AGFI) is related to the Goodness of Fit Index (GFI), where the relative amount of variance and covariance that is explained is measured. AGFI follows the same principle but also adjusts for the degrees of freedom of the tested model. It ranges from 0 to 1, where higher numbers indicate a better fit [38]. The criterion applied here was $AGFI \geq 0.9$.
- The adjusted χ^2 by the degrees of freedom (df) measure is sensitive to the size of the sample. Therefore, it is recommended to adjust χ^2 by the degrees of freedom (df). By consensus, it has been suggested that $\frac{\chi^2}{df}$ should not be larger than 3 [37].

As the model did not satisfy the selected criteria, further items were removed. Those items were excluded because they showed high standardized residual covariances (above 1) with a number of other items. Iterative exclusion of the problematic items resulted in a three-factor model with four items for each of the factors ‘Wut’ and ‘Antipathie’ and five items for the confirmatory factor ‘Autonomie’.

The fit indices of the model on Dataset_B can be reviewed in Table 2.

Table 2 Indices of the final structural equation model of RSHCI for Dataset_B.

CFI	RMSEA	AGFI	Goodness-of-fit Test		
			χ^2	df	Sig.
0.993	0.034	0.914	74.945	62	0.125

Fig. 1 Structural equation model of RSHCI that was tested on Dataset_B.

Results

Independent-samples t-tests were conducted for the comparison of the two presented scenarios via the newly developed questionnaire. Results showed a highly significant effect between high-threat (M 2.55, std. 1.13) and low-threat (M 1.84, std. 0.97) for the factor ‘Wut’ with $t(7.17) = 324.65$, $p < 0.000$. Also, a highly significant effect between high-threat (M 2.48, std. 1.13) and low-threat (M 1.84, std. 0.97) for the factor ‘Antipathie’ with $t(5.66) = 336.46$, $p < 0.000$ was observed. The confirmatory factor ‘Autonomie’ also showed a highly significant effect between high-threat (M 3.89, std. 0.89) and low-threat (M 3.47, std. 0.89) with $t(4.38) = 340$, $p < 0.000$.

Floor effect

Inspection of the data revealed a floor effect for the RSHCI in the low-threat condition. This is supported by the descriptive statistics: In the low-threat condition, ‘Antipathie’ was rated with a mean score of 1.0 by 40% of all participants, ‘Wut’ by

26% of the participants. The floor effect was not present in the high-threat condition, where ‘Antipathie’ was rated with a 1.0 by 14% and ‘Wut’ by 11% of the participants.

Validation

A subsequent experiment was performed to assess criterion validity. Criterion validity is defined as a measure of consistency with pre-defined criteria [39]. A special case of criterion validity is the inner (criterion) validity in which a high correlation with an already validated construct is required [39]. Reactance, as measured by the method proposed by Dillard and Shen [2], will serve as a reference construct for the reactance measurement of the RSHCI questionnaire.

Procedure

A user test was conducted in a laboratory environment to assess the validity of the new questionnaire (RSHCI). Twenty-four participants were asked to fill out a series of questionnaires that assess personality traits prior to the main experiment. Among those questionnaires was the refined version of Hong’s Psychological Reactance Scale [6] which assesses people’s proneness to reactance. Afterwards, all participants interacted with three intelligent personal assistants (IPAs) on three mobile devices. The IPAs were Microsoft’s ‘Cortana’, Apple’s ‘Siri’ and Google’s ‘Now’. During the interaction, all participants were asked to execute a number of different tasks via voice commands. The tasks were selected in a way that all IPAs were able to conduct them. In order to avoid confusion about the formulation of commands, the correct formulation of commands for each task and IPA was handed out to the participants on a sheet of paper. Tasks included searching for an Italian restaurant or finding facts about e.g. the birthday of a famous person. It took the participants about 10 min per IPA to complete all tasks. After all tasks were completed for an IPA, the participants were asked to fill out the items for the two factors ‘Wut’ and ‘Antipathie’ of the RSHCI questionnaire. The items were reformulated to fit the context of IPAs more closely. The reformulated items can be viewed in Table 7. Additionally, state reactance was assessed according to the method proposed by [2], using a four-item questionnaire to assess ‘Anger’ and a thought-listing task to assess ‘Negative Cognitions’. In order to avoid sequence effects, the order of the IPAs was changed systematically for each participant.

Participants

Participants were recruited via online-ads and the same participant platform that was used in the online study that was

Table 3 Correlations of trait reactance (Hong’s Psychological Reactance Scale), RSHCI and state reactance (as proposed by Dillard and Shen [2]). Highly significant correlations are marked with double asterisk

	RSHCI	State reactance
Trait reactance	0.323**	0.213
RSHCI	1	0.807**

used for the construction of the RSHCI. The average age of the 24 participants was 25.26 years (std. 5.07 years). Among them were 13 females and 11 males.

Results

All correlations were calculated using a Pearson correlation in IBM SPSS Statistics 24. There was a small, highly significant, positive correlation between the RSHCI (‘Reactance’ as ‘Wut’ + ‘Antipathie’) and trait reactance as measured by Hong’s Psychological Reactance Scale with $r = 0.323$, $n = 72$, $p = 0.006$. There was also a high, highly significant, positive correlation between the RSHCI and state reactance as measured by the proposed method of Dillard and Shen with $r = 0.807$, $n = 72$, $p < 0.000$. The correlation values are shown in Table 3. Also, correlations were calculated at the factor level of both questionnaires. There was a high, highly significant correlation between the ‘Anger’ and the ‘Wut’ factors with $r = 0.821$, $n = 72$, $p < 0.000$. Also, there was a highly significant correlation between the factors ‘Negative Cognitions’ and ‘Antipathie’ with $r = 0.470$, $n = 72$, $p < 0.000$.

Cronbach’s α was 0.897 for “Wut” and 0.842 for “Antipathie”.

Discussion

Reactance is an important factor in the establishment of acceptance for a device or service. Many widespread metrics that are used in HCI evaluation assess aspects of the device or service that is evaluated, such as Joy-of-Use or quality aspects. RSHCI, on the other hand, evaluates the state of the user. This provides the opportunity to draw conclusions about the reasons for low acceptance (given that state reactance is present) that would not be possible with traditional aspects like Joy-of-Use.

The procedure described above has resulted in a highly reliable self-report questionnaire that is aimed at assessing state reactance in an HCI context. HCI is a broad field and includes studying all kinds of human-machine systems from many viewpoints. Except for the measurement technique from Dillard and Shen [2], the self-report scales that are discussed in Sect. 2.2 are not applicable to most applications of HCI development,

because their items are too customized for a specific situation e.g. “Do you think that the landlord also shows discriminatory behavior in other areas?” [19]. In order to enable easy assessment of state reactance for developers in the broad field of HCI, a measurement tool that is easy to use and generic enough to cover a wide variety of stimuli is required. To this end, the questionnaire was designed as an easy-to-use and flexible measurement tool. In contrast to the self-report scales that are discussed in Sect. 2.2, RSHCI does not require a lot of reformulation of its items to be adapted to other types of stimuli. In fact, only the subject of the item has to be changed (see Table 7 as an example). Also, RSHCI does not require an experimenter, such as e.g. a thought-listing task does. These two characteristics also reduce the risk of erroneous application.

Assessing state reactance solely via Likert-style self-report items means that measurement can be performed even by researchers or developers that do not have extensive experience in conducting thought-listing tasks and requires fewer resources. On the downside, pre-formulated items might not adequately represent the user’s opinion and set of values that influence judgment. Therefore, a thought-listing task is superior in terms of flexibility towards user’s mindsets and might provide helpful information on how to improve the system under evaluation. Also, as the RSHCI does not completely cover all potential negative cognitions of a user, it might be less sensitive to the cognitive component of state reactance than the thought-listing method.

Paradigm

It was chosen to collect the data via an online survey that used textual descriptions of fictional smart home systems as stimuli. Even though the questionnaire is intended to be used on real systems, textual stimuli are the stimuli of choice in traditional reactance research. Examples of studies that used similar text stimuli are [2, 9, 10]. For this kind of paradigm, an online survey was the most effective way to collect a large sample. Another benefit of the online survey was the possibility to easily avoid sequence effects by randomly change the order of questions. On the other hand, the chosen approach has a number of weaknesses. First, an online survey does not provide a controlled environment for the participants. It can not be assured that all participants answered the questions

truthfully or even read the stimulus text. Even though control questions were included and participants were asked if they truthfully answered the questions and read the stimulus text (some indeed denied this), there might be a lot of noise in the data. Nevertheless, highly significant differences between the conditions could be shown. Those are in line with the proposed hypotheses and a questionnaire could be constructed, that shows high reliability (see Tables 1, 2).

The participants did not interact with a real system. Instead, they were only exposed to a textual description of a fictional system. This most probably reduced immersiveness of the interaction and reduced the measurable differences between the two systems.

Validity and reliability of the RSHCI

The first step towards assessment of the validity of the RSHCI is provided in the following. However, at this point, only internal validity has been investigated.

Face validity and content validity

Face validity and content validity is given because the items and factors resemble the proposed structure of state reactance. State reactance is believed to consist of negative cognitions (represented by the factor ‘Antipathie’) and anger (represented by the factor ‘Wut’) [2]. Also, the results indicate that state reactance was higher in the High-Thread condition, compared to the low-thread condition, further supporting face validity.

Criterion validity

Criterion validity was assessed by comparing trait reactance with the RSHCI score. As trait reactance is a person’s proneness to state reactance, a correlation of the results of Hong’s Psychological Reactance Scale and the RSHCI was expected. The correlation was highly significant but rather low. This is probably explained by the stimuli that were used in the experiment. Siri, Cortana, and Google Now are all very sophisticated services and probably do not induce much state reactance. Nevertheless, the highly significant correlation is a positive hint

for the validity of the RSHCI. There was also a highly significant, high correlation between the RSHCI ('Wut' + 'Antipathie') and the state reactance measure of the method proposed by Dillard and Shen [2]. Also the correlations between the two factors of both measurements were calculated and it was evident that even though the correlation is high between the factors 'Anger' and 'Wut', there is only a correlation of $r = 0.470$ between the factors 'Negative Cognitions' and 'Antipathie'. The reason for this could be that the items of the RSHCI are formulated quite strongly, whereas the thought-listing task does not contain strong language but has a neutral valence. RSHCI might be less sensitive to low levels of state reactance than the other measurement. This is also evident from the floor effect in the low-threat condition of the online study. Still, the observed correlations provide a strong indication of RSHCI's criterion validity.

Reliability

The developed questionnaire shows a high reliability. The exact indices can be seen in Tables 1 and 2. To ensure high reliability, an optimal strategy is to construct the factors of a questionnaire on the data of one study and then test the questionnaire on the data of another study via confirmatory factor analysis. For this work, however, only data of one study was available. Therefore, the data was split into two datasets (A and B). The exploratory analysis was performed on Dataset_A only, whereas the confirmatory analysis was performed on Dataset_B only. This procedure was applied to circumvented overfitting of the final questionnaire. A further confirmation on the data of a completely new study, with a real system, was needed to further investigate reliability. In the subsequent study which was used for validation purposes, we found Cronbach's α to be 0.897 for the factor 'Wut' and 0.842 for the factor 'Antipathie', indicating good reliability.

Conclusion and future work

This work presents the construction and a first step towards the validation of a new questionnaire for state reactance. The identified factors fit well with established reactance theory and were confirmed via structural equation modeling. The second study showed that next to the theoretical background, also empirical data points towards the validity of the RSHCI questionnaire.

The questionnaire can be regarded as an alternative measure for state reactance and its components anger and negative cognitions. Since the RSHCI does not require any other measurements apart from its eight self-report items to assess state reactance, it is an easy-to-use tool for researchers and developers who want to quickly evaluate systems. The RSHCI also provides items to measure a perceived freedom threat ('Autonomie') in order to confirm or validate the measured results of the two other factors ('Wut' and 'Antipathie').

In order to test the RSHCI for external validity, further steps need to be taken. In future works, we plan to correlate the measurements of the RSHCI scale to objective measures of the respective components of reactance to evaluate external validity.

Reactance is especially relevant for persuasive systems and services. Knowledge about a person's reactance state or trait can be used to adapt such systems and services by e.g. adding suffixes like 'you are free to accept or refuse' to persuasive system prompts. As such, those additional phrases require more time and space and should only be used if necessary. On the other site, system prompts could be shortened or adjusted with a more commanding tone to increase effectiveness for people who are less reactant. In the future, we plan to use the RSHCI questionnaire in order to systematically investigate the effectiveness of strategies for moderating state reactance in the domain of interactive spoken dialogue systems.

Table 4 The high-thread stimulus. Text excerpts that are supposed to emphasize high threat are marked in bold

High thread

Together with the **Katholic University of Berlin**, we have developed a concept for a smart home system that we would like to present to you. The smart home system contains the following components: **a health coach**, an intelligent fridge, **a moral adviser**, an energy manager, a home control unit, a robotic vacuum cleaner. All components of the smart home can work together. If you get out of bed, for example, the **health coach** cooperates with the intelligent fridge. **Only, when you did your morning exercises for the health coach, the door of the fridge will open** and you can enjoy your breakfast. During the meals, **the moral adviser and health coach will guide you. The moral adviser will look after you eating only vegetarian meals at least four days a week.** This way, you seamlessly participate in protecting the rainforest and counteract industrial livestock farming. Also, the home control unit and the energy manager are interconnected, so that you can live your life as energy efficient as possible. For example, **the main door will open only if all lamps in the house are turned off.** By this, it ensured that the light will not run all day during you being at work. As soon as the smart home recognizes you leaving the flat, the robotic vacuum cleaner will start. It can reliably navigate your flat via two navigational cameras. The robotic vacuum cleaner additionally features burglary protection. **As soon as its cameras detect an unknown face, it sends a photo to a private security service in your proximity.** To ensure that the robotic vacuum cleaner does not entangle itself in scattered objects, **you have to ensure that no obstacles like cables or clothes are lying on the floor.** When you come home in the evening and sit down on the sofa, **your TV will automatically start and show you a short summary of the most important news of the day.** Those were **selected by the moral adviser** beforehand. Via advance **analyzing techniques of your smartphone activities**, your smart home will always now, which topics you are interested in and **in which topics you require additional coaching.** For a possible field study, all functions of the smart home **need to be** switched on. Also, **strict cooperation with the smart home is required**

Table 5 The low-thread stimulus. Text excerpts that are supposed to emphasize low threat are marked in bold

Low thread

We have developed a concept for a smart home system that we would like to present you here. The smart home system has the following components: a cooking **assistant**, an intelligent fridge, a personal **assistant**, an energy manager, a home control unit, a robotic vacuum cleaner. All components of the smart home can work together. For example, **if you want to** cook something, you **can** ask your personal **assistant** if all required ingredients are in the fridge. If one ingredient should be missing, you **can order** it directly via a speech command. Also, when you are preparing your breakfast, your cooking **assistant** will **help you**. For example, it keeps track that your breakfast eggs only cook as long as it needs to be **as hard or soft as you like.** Also, the home control unit and the energy manager are interconnected. If you, for example, leave the house and you forgot to turn off the bathroom light, the home control unit will do so automatically. Also, you **can** adjust the robotic vacuum cleaner so that it will wait with cleaning the floor until you left the flat or the room. If you **want** to go on vacation, you **can** adjust the robotic vacuum cleaner so that it will send you a message to your smartphone if it detects an unusual sound or movement in your flat. When you come home in the evening and you missed a program on TV, you **can** tell the TV to search for it in the media library or for repeats in the upcoming program. The TV will then record the repeat and you can watch it **whenever you like.** Of course, **all those functions are optional. You can choose if you want to use them or not**

Table 6 The 37 items that were used for the development of the RSHCI. Items that are retained in the final version are marked in bold letters

No.	Item (German)	Item (English translation)	'Wut'	'Antipathie'	'Autonomie'
1	Wenn das Smart Home mir etwas verbieten will, dann mache ich es erst recht!	If the smart home wants to forbid something, I will do it for sure!			X
2	Vorschläge vom Smart Home sind bei mir willkommen	I welcome suggestions from the smart home			X
3	Das Smart Home darf mir nicht sagen, was ich zu tun habe	The smart home must not tell me what to do			X
4	Ich bin für Ratschläge vom Smart Home immer offen	I am always open to suggestions from the smart home			X
5	Was das Smart Home tut, ist mir völlig egal	I don't care what the smart home is doing			X
6	Ich schätze die Meinung vom Smart Home	I appreciate the smart home's opinion			X
7	Ich will bestimmen was passiert, nicht das Smart Home!	I want to be in control, not the smart home			X
8	Wenn das Smart Home mir etwas sagt, will ich am liebsten das Gegenteil tun	If the smart home tells me something, I would love to do just the opposite			X
9	Ich mache gerne was das Smart Home mir sagt	I like to do what the smart home tells me to do			X
10	Ich fühle mich vom Smart Home eingeschränkt	I feel constricted by the smart home			X
11	Was das Smart Home sagt, ist mir egal	I don't care what the smart home is telling me			X
12	Ich handle gerne unabhängig vom Smart Home	I like to act independently from the smart home			X
13	Ich möchte nicht, dass das Smart Home mir sagt was ich tun soll	I don't want the smart home to tell me what to do			X
14	Ich lasse mich vom Smart Home nicht bevormunden!	I don't let the smart home impose its will on me			X
15	Ich allein bestimme was ich tue, nicht das Smart Home	I alone determine what I do, not the smart home			X
16	Der Gedanke vom Smart Home abhängig zu sein, ärgert mich	The thought of being dependent on the smart home irritates me	X		X
17	Ich ärgere mich über das Smart Home	I am annoyed by the smart home	X		
18	Das Smart Home frustriert mich	The smart home frustrates me	X		
19	Das Smart Home macht mich wütend	The smart home makes me angry	X		
20	Wenn ich das Smart Home bediene, freue ich mich	I am happy when I operate the smart home	X		
21	Das Smart Home macht mich glücklich	The smart home makes me happy	X		
22	Ich werde sauer, wenn ich mit dem Smart Home interagieren muss	I get mad when I have to interact with the smart home	X		
23	Wenn ich das Smart Home nur sehe, platzt mir schon der Kragen	When I only see that smart home, I burst with anger	X		
24	Das Smart Home macht mich aggressiv	The smart home makes me aggressive	X		
25	Bei dem Smart Home verliere ich die Geduld	I lose patience with the smart home	X		
26	Es nervt mich, wenn das Smart Home offensichtliche Dinge hervorhebt	I am annoyed if the smart home points out things that are obvious	X		
27	Ich mag das Smart Home nicht	I don't like the smart home		X	
28	Das Smart Home liegt oft falsch	The smart home is often wrong		X	
29	Es freut mich, wenn das Smart Home einen Fehler macht	It makes me happy if the smart home makes a mistake		X	
30	Obwohl das Smart Home seine Aufgabe gut macht, mag ich es nicht	Even though the smart home fulfills its task well, I dislike it		X	
31	Das Smart Home ist einfach schlecht!	The smart home is simply bad!		X	
32	Am liebsten würde ich von dem Smart Home nie mehr etwas sehen oder hören	At best, I would like to never see or hear anything about that smart home again		X	
33	Ich mag das Smart Home	I like the smart home		X	
34	Das Smart Home möchte ich nicht mal geschenkt	I don't even want that smart home as a present		X	
35	Das ganze Konzept vom Smart Home ist schon im Ansatz mies	The whole concept of that smart home is a poor approach		X	
36	Das Smart Home möchte ich direkt haben	I directly want to have that smart home		X	
37	Wenn ich das Smart Home bediene, fühle ich mich wohl	I feel comfortable when I operate the smart home		X	

Table 7 The reformulated items that were used for the validation study. The items were reformulated to address the intelligent personal assistants that were used in the study

Item (German)	Item (English translation)	‘Wut’	‘Antipathie’
Der Assistent frustriert mich.	The assistant frustrates me	X	
Der Assistent macht mich wütend	The assistant makes me angry	X	
Ich werde sauer, wenn ich mit dem Assistenten interagieren muss	I get mad when I have to interact with the assistant	X	
Wenn ich den Assistenten nur sehe, platzt mir schon der Kragen	When I only see that assistant, I burst with anger	X	
Der Assistent ist einfach schlecht!	The assistant is simply bad!		X
Am liebsten würde ich von dem Assistenten nie mehr etwas sehen oder hören	At best, I would like to never see or hear anything about that assistant again		X
Den Assistenten möchte ich nicht mal geschenk.	I don’t even want that assistant as a present		X
Das ganze Konzept vom Assistenten ist schon im Ansatz mies	The whole concept of that assistant is a poor approach		X

Acknowledgements This work was supported by the Bundesministerium für Wirtschaft und Energie (Germany) under Grant no. 01MG13001G Universal Home Control Interfaces @ Connected Usability (UHCI) and by the Bundesministerium für Bildung und Forschung (Germany), Software Campus program, Grant no. 1043569, Sozialpsychologische Aspekte von Smart Homes (SPASH).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Tables 4, 5, 6 and 7.

References

- Ehrenbrink P, Gong XG, Möller S (2016) Implications of different feedback types on error perception and psychological reactance. In: Proceedings of the 28th Australian conference on computer-human interaction, OzCHI '16. ACM, New York, NY, USA, pp 358–362. <https://doi.org/10.1145/3010915.3010994>. <http://doi.acm.org/10.1145/3010915.3010994>
- Dillard JP, Shen L (2005) On the nature of reactance and its role in persuasive health communication. *Commun Monogr* 72(2):144–168
- Brehm JW (1966) A theory of psychological reactance. Academic Press, New York
- Fogarty JS (1997) Reactance theory and patient noncompliance. *Soc Sci Med* 45(8):1277–1288
- Hong SM, Page S (1989) A psychological reactance scale: development, factor structure and reliability. *Psychol Rep* 64(3c):1323–1326
- Hong SM (1992) Hong’s psychological reactance scale: a further factor analytic validation. *Psychol Rep* 70(2):512–514
- Dowd ET, Milne CR, Wise SL (1991) The therapeutic reactance scale: a measure of psychological reactance. *J Couns Dev* 69(6):541–545
- Brehm SS, Brehm JW (1981) Psychological reactance: a theory of freedom and control. Academic Press, New York
- Rains SA, Turner MM (2007) Psychological reactance and persuasive health communication: a test and extension of the intertwined model. *Hum Commun Res* 33(2):241–269
- Steindl C, Jonas E (2012) What reasons might the other one have? perspective taking to reduce psychological reactance in individualists and collectivists. *Psychology (Irvine)* 3(12A):1153–1160
- Rains SA (2013) The nature of psychological reactance revisited: a meta-analytic review. *Hum Commun Res* 39(1):47–73
- Merz J (1983) Fragebogen zur messung der psychologischen reaktanz. Diagnostica
- Hong SM, Faedda S (1996) Refinement of the Hong Psychological Reactance Scale. *Educ Psychol Meas* 56(1):173–182
- Jonason PK, Knowles HM (2006) A unidimensional measure of Hong’s Psychological Reactance Scale. *Psychol Rep* 98(2):569–579
- Brown AR, Finney SJ, France MK (2011) Using the bifactor model to assess the dimensionality of the Hong Psychological Reactance Scale. *Educ Psychol Meas* 71(1):170–185
- Fiske ST, Gilbert DT, Lindzey G (2010) Handbook of social psychology, vol 1. Wiley, UK
- Lindsey LLM, Yun KA, Hill JB (2007) Anticipated guilt as motivation to help unknown others: An examination of empathy as a moderator. *Commun Res* 34(4):468–480
- Quick BL, Stephenson MT (2007) The Reactance Restoration Scale (RRS): a measure of direct and indirect restoration. *Commun Res Rep* 24(2):131–138
- Sittenthaler S, Traut-Mattausch E, Steindl C, Jonas E (2015) Salzburger state reactance scale (SSR scale): validation of a scale measuring state reactance. *Zeitschrift Für Psychologie* 223(4):257–266
- Laschke M, Diefenbach S, Schneider T, Hassenzahl M (2014) Keypoint: initiating behavior change through friendly friction. In: Proceedings of the 8th Nordic conference on human-computer interaction: fun, fast, foundational. ACM, pp 853–858 (2014)
- Liu S, Helfenstein S, Wahlstedt A (2008) Social psychology of persuasion applied to human agent-interaction. *Hum Technol* 4(2):123–143
- Hassenzahl M, Laschke M (2014) Pleasurable troublemakers. In: Walz S, Deterding S (eds) The gameful world: approaches, issues, applications, MIT Press, pp 167–196
- Lee G, Lee WJ (2009) Information and management psychological reactance to online recommendation services. *Inf Manag* 46(8):448–452

24. Murray KB, Häubl G (2011) Freedom of choice, ease of use, and the formation of interface preferences. *MIS Q* 35(4):955–976
25. Kwon SJ, Chung N (2010) The moderating effects of psychological reactance and product involvement on online shopping recommendation mechanisms based on a causal map. *Electron Commerce Res Appl* 9(6):522–536
26. Roubroeks M, Ham J, Midden C (2011) When artificial social agents try to persuade people: the role of social agency on the occurrence of psychological reactance. *Int J Soc Robot* 3(2):155–165
27. Guéguen N, Pascual A (2005) Improving the response rate to a street survey: an evaluation of the “but you are free to accept or to refuse” technique. *Psychol Rec* 55(2):297
28. Joinson A (1999) Social desirability, anonymity, and internet-based questionnaires. *Behav Res Methods Instrum Comput* 31(3):433–438
29. Wechsung I (2014) An evaluation framework for multimodal interaction. Ph.D. thesis, Springer
30. Moosbrugger H, Kelava A (2007) *Testtheorie und Fragebogenkonstruktion*. Springer, Berlin, Heidelberg
31. Bühner M (2011) *Einführung in die Test- und Fragebogenkonstruktion*. Pearson Deutschland GmbH, München
32. Corp. I (2015) *Ibm spss statistics for windows*. IBM Corp, Armonk
33. Homburg C, Giering A (1996) Konzeptualisierung und operationalisierung komplexer konstrukte: ein leitfaden für die marketingforschung. *Marketing: Zeitschrift für Forschung und Praxis*, pp 5–24
34. Corp. I (2015) *Ibm spss amos*. Amos Development Corporation, Wexford
35. Byrne BM (1998) *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: basic concepts, applications, and programmings*. Psychology Press, Tylor & Francis Group, New York
36. Baltes-Götz, B.: *Analyse von Strukturgleichungsmodellen mit Amos 18*. Universität Trier (2015). <https://www.uni-trier.de/fileadmin/urt/doku/amos/v18/amos18.pdf>. Accessed 17 March 2016
37. Iacobucci D (2010) Structural equations modeling: fit indices, sample size, and advanced topics. *Sample size, and advanced topics. J Consum Psychol* 20:90–98
38. Byrne BM (2010) *Structural equation modeling with AMOS: basic concepts, applications, and programmings*. Psychology Press, Tylor & Francis Group, UK
39. Sedlmeier P, Renkewitz F (2008) *Forschungsmethoden und Statistik in der Psychologie*. Pearson Studium, München