

## University of Groningen

### Public service guarantees

Thomassen, Jean Pierre Robert

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2018

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Thomassen, J. P. R. (2018). Public service guarantees: Exploring the design and implementation of service guarantees in public settings. [Groningen]: University of Groningen, SOM research school.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

## **Chapter 3.**

### **Compensating customers for poor service delivery: Experimental research in public and private settings<sup>10</sup>**

---

<sup>10</sup> This chapter has been published in *Public Administration* (IF: 2.959; AI 0.76) in December 2017 (95(4), 895-911); co-authors are Dr. M.C. Leliveld, Prof. S. Van de Walle and Prof. C.T.B. Ahaus. Together with my three co-authors I developed the experiments. In close cooperation with the second author I did the data gathering and analysed the results. I have written the paper, my three co-authors have reviewed the paper and helped improving it.

### 3.1. Introduction

A driving license that is not available at the agreed date or a customer who has waited too long at a counter are both examples of operational service failures. They are inevitable and part of daily life (Kim and Ulgado, 2012) because services are intangible, hard to standardise and production and consumption happen simultaneously (Murray and Schlachter, 1990). They are little studied in public management literature and how customers recover satisfaction after a service failure (service recovery) even less (Van de Walle, 2016). This chapter uses service management literature from the private domain as its theoretical base. Here, justice theory (Adams, 1965) is the dominant framework to explain customers' evaluations and behaviour after a service failure. It argues that customers evaluate the fairness of service recovery on distributive, procedural and interactional justice (e.g. Homburg and Fürst, 2005; Vázquez-Casielles *et al.*, 2010).

Expectancy disconfirmation theory (Oliver, 1993) states that differences between expectations and experiences (e.g. for justice dimensions) can lead to positive or negative disconfirmations, which subsequently influence post-recovery satisfaction (e.g. Van Ryzin, 2013). Perceived justice and satisfaction after a service failure are influenced by how organisations operate after a failure; for example, by apologising, fixing the problem and/or offering compensation. Monetary compensation is a common recovery instrument to improve perceived justice and satisfaction after a service failure (e.g. Wirtz and Mattila, 2004). Compensation schemes exist in the private domain (e.g. hotels and airlines) and the semi-public domain (e.g., railways, city transport and energy supply) but are less common in core public organisations. Public organisations also differ in whether they explicitly promise to compensate when a service failure occurs.

To understand public compensation, I researched the effects of promising and offering a small monetary compensation on customers' evaluations of perceived justice, negative emotions and post-recovery satisfaction in public and private settings. I used a between-subjects factorial design across Dutch students (study 1) and replicated this study using a US online panel (study 2) to increase external validity. The results contribute to theory in two ways. First, by adding to the literature on the research theme. Second, this is the first experimental study to apply justice theory in public service recovery settings. Justice theory emerges as applicable to public as in private settings. The next sections elaborate by reviewing current empirical service management literature on service failures, service recovery, justice theory

and monetary compensation to formulate hypotheses. Then the experimental methodology and results are presented. I conclude this chapter by discussing limitations and finally suggest future research.

### **3.2. Service failures**

New Public Management ideas suggest public organisations increasingly approach citizens as customers (Aberbach and Christensen, 2005). Customer satisfaction metrics have become important for public managers (Van Ryzin, 2013). Therefore, public management scholars (Osborne *et al.*, 2015) argue for a service-dominant logic approach (Lusch and Vargo, 2014) to place customers, rather than products, policy makers or professionals, at the heart of service research, design and operations. From this perspective, service failures are defined as situations in which customers experience an economic (e.g., money, time) and/or a social loss (e.g. status, esteem) due to a mishap or a problem when experiencing a public service (Kim and Ulgado, 2012) regardless of responsibility (Magnini *et al.*, 2007). Marketing and services management literature categorise failures by their type and severity. First, there are process and outcome failures (Tsai *et al.*, 2014). Process failures occur during service delivery and involve how customers receive the service. Whereas, outcome failures involve what customers actually receive. Outcome failures include delay versus denial failures. A delay requires customers to wait to receive service, whilst denial is the total breach of a (implicit) contract (Levesque and McDougall, 2000). Second, the severity or importance of a failure depends on customers' perceived cumulated economic and social loss resulting from service failure. Failures range from unimportant and mildly annoying through to extremely important and very severe (Magnini *et al.*, 2007; Mattila, 2001). The type of failure and the severity both affect customers' perceived loss and unfairness.

### **3.3. Service recovery and justice theory**

Service recovery involves actions organisations take to respond to service failure to make up for the perceived loss sufficient to regain customers' satisfaction (Hocutt and Bowers, 2005; Mattila, 2001). The larger the loss customers feel, the more recovery they seek (Kim and Ulgado, 2012). Justice theory (Adams, 1965) sees customers evaluating recovery fairness in interactional, procedural and distributive terms. Interactional justice is the perceived fairness of treatment by employees. Procedural justice is the perceived fairness of the organisation's recovery policies and processes. Distributive justice is the perceived fairness of the outcome, such as a monetary compensation. Many studies show that, depending on the service and

service failure context, the relative impact of these three justice dimensions on post-recovery satisfaction is different (Del Río-Lanza *et al.*, 2008; Mattila, 2001).

Expectancy disconfirmation theory (Oliver, 1993; Van Ryzin, 2013) sees customer satisfaction being determined by the difference between customers' experiences and their expectations. This is relevant for service recovery experiences and expectations resulting in post-recovery satisfaction. For most situations customers have expectations, these are norms and standards or benchmarks against which customers judge or measure the quality of service they receive (Magnini *et al.*, 2007). Expectations are influenced by factors such as past experiences, word-of-mouth and communication by the organisation (Zeithaml and Bitner, 1996). Expectations are 'formed in a rich context of remembered and constructed representations of what it could have been, might have been, or should have been' (Kahneman and Miller, 1986, p. 136). In literature, expectations are divided into positive and normative expectations (e.g. James, 2011; Yim *et al.*, 2003). Positive expectations are customers' predictive norms concerning what actually will happen. Normative expectations are based on what should happen according to the customer, for example concerning service recovery after a service failure.

#### **3.4. Monetary compensation for poor service**

Monetary compensation is a financial value customers receive in order to (partly) balance the perceived economic and/or social loss due to a service failure. Authorities and governments have obliged semi-public organisations to offer monetary compensation for poor service. For example, railways and public transport offer compensation schemes for punctuality and reliability (Björölin Lidén and Edvardsson, 2003) and energy supply companies for outage and other service failures (Costello, 2012). While monetary compensation is common in semi-public and private organisations, it is rare for core public organisations given ethical and legal arguments. Ethically, as public organisations are financed by taxpayer money, this should be spent on the collective and not on individual customers. Also, compensating individuals may increase inequality in service delivery between customers (Fountain, 2001; Van de Walle, 2016). Legally, national legislation could influence the possibilities to offer compensation. As in private settings, customers can expect quality service without failures and value for the money when they pay indirectly through taxes or directly through fees for services. When failure occurs, compensation helps balance the loss. Thus, it makes sense for public organisations to offer compensation for poor service.

Like service failures, monetary compensation can be categorised by type and size. The type and size have to match the type (Roschk and Gelbrich, 2014) and severity (Gelbrich *et al.*, 2015) of the failure to be effective. Monetary types of compensation include gift vouchers and coupons, discounts, money back or free products and services (e.g. Lii and Lee, 2012). Compensation is offered proactively or reactively after a complaint and on the spot at the time of the failure or delayed (Kim and Ulgado, 2012). The compensation size varies from no compensation, a small and token compensation not directly related to nor fully compensating for the loss, to equity compensation equal to the loss or even overcompensating customers. Research in private settings shows that compensation size has effects on customers' evaluations (Haesevoets *et al.*, 2017; Hocutt and Bowers, 2005). Offering a small compensation that is given as a token or gesture to customers has different effects on customers' evaluations than fully compensating or overcompensating customers (Gelbrich *et al.*, 2015; McQuilken *et al.*, 2013). In my studies I researched the effects of a small compensation after a service failure because being overly generous following a service failure may lead customers to question the reasons behind offering compensation (McQuilken *et al.*, 2013) and having possible problems with spending large sums of public money on compensating for failures (Björölin Lidén and Edvardsson, 2003).

What exactly happens when customers experience service failure? First, there is customers' cognitive appraisal of justice dimensions comparing experiences with expectations. The perceived justice elicits a negative emotional reaction (Schoefer and Ennew, 2005) such as regret, annoyance, irritation, anger and feeling betrayed (Mattila, 2001). Both perceived justice and negative emotions impact post-recovery satisfaction. Research shows compensation can have positive effects on post-recovery satisfaction via the increase in perceived distributive justice (e.g. Schoefer and Ennew, 2005; Wirtz and Mattila, 2004) and the reduction of negative emotions (Del Rio-Lanza *et al.*, 2008). The specifics of how compensation is offered relates to procedural aspects which we also study. Previous literature suggests that procedural justice has a positive effect on post-recovery satisfaction (Vázquez-Casielles *et al.*, 2010).

Justice theory has yet to be applied in public service recovery research. However, I believe it is highly relevant for public management research; especially in direct exchange situations when public customers pay directly for the service (Alford, 2002). The first hypothesis (H1)

about offering compensation after service failures in direct exchange situations is: perceived justice, emotions and post-recovery satisfaction are more positive when compensation is offered (Comp) compared to when it is not offered (NoComp) after a service failure.

Note: to help interpretation in the results sections I use abbreviations of cell names as presented in the hypotheses; Prom (a compensation is promised) versus NoProm (no compensation is promised) and Comp (a compensation is offered) versus NoComp (no compensation is offered).

### **3.5. Promising compensation**

This research also focuses on the effects of explicitly promising compensation prior to a service failure by using service guarantees (see Chapter 2 for an overview of the concept). When explicitly promised, compensation acts as a cue increasing customers' expectations when compared with situations where no compensation is promised. According to expectancy disconfirmation theory (Oliver, 1993) I could argue that promising compensation has negative effects on customers' evaluations. This leads to H2: in situations where compensation is promised explicitly and offered (Prom-Comp), perceived justice, emotions and post-recovery satisfaction are more negative than when compensation has not been promised but is offered (NoProm-Comp).

In H1 I hypothesised that compensating customers would lead to more positive evaluations. In H2 I expect that when compensation is offered, promising compensation leads to less positive evaluations. When combining these hypotheses, the question is whether offering compensation (promised or not-promised) leads to more positive evaluations than when compensation is neither promised nor offered. I hypothesise in H3: when compensation is offered, either with (Prom-Comp) or without prior promise (NoProm-Comp), perceived justice, emotions and post-recovery satisfaction are more positive compared to when compensation is neither promised nor offered (NoProm-NoComp).

### **3.6. Double deviation: failures in service recovery**

Making promises creates a risk when you cannot keep them. Marketing and services management research has shown that more than half of all attempted recovery efforts reinforce dissatisfaction because of failed service recoveries (Casado-Díaz and Nicolau-Gonzálbez, 2009). The expectancy disconfirmation theory explains this so-called double

deviation effect. Explicitly promising compensation increases customers' expectations. Not keeping this promise leads to extremely low levels of customers' evaluation and their criticism damaging the organisation (Casado-Díaz and Nicolau-Gonzálbez, 2009). Hypothesis 4a is: when compensation is promised, but not offered (Prom-NoComp), perceived justice, emotions and post-recovery satisfaction are more negative compared to when compensation is not promised nor offered (NoProm-NoComp). The related hypothesis 4b is: when compensation is promised, but not offered (Prom-NoComp), perceived justice, emotions and post-recovery satisfaction are more negative compared to when compensation is offered, both promised (Prom-Comp) and not promised (NoProm-Comp).

### **3.7. Public - private differences**

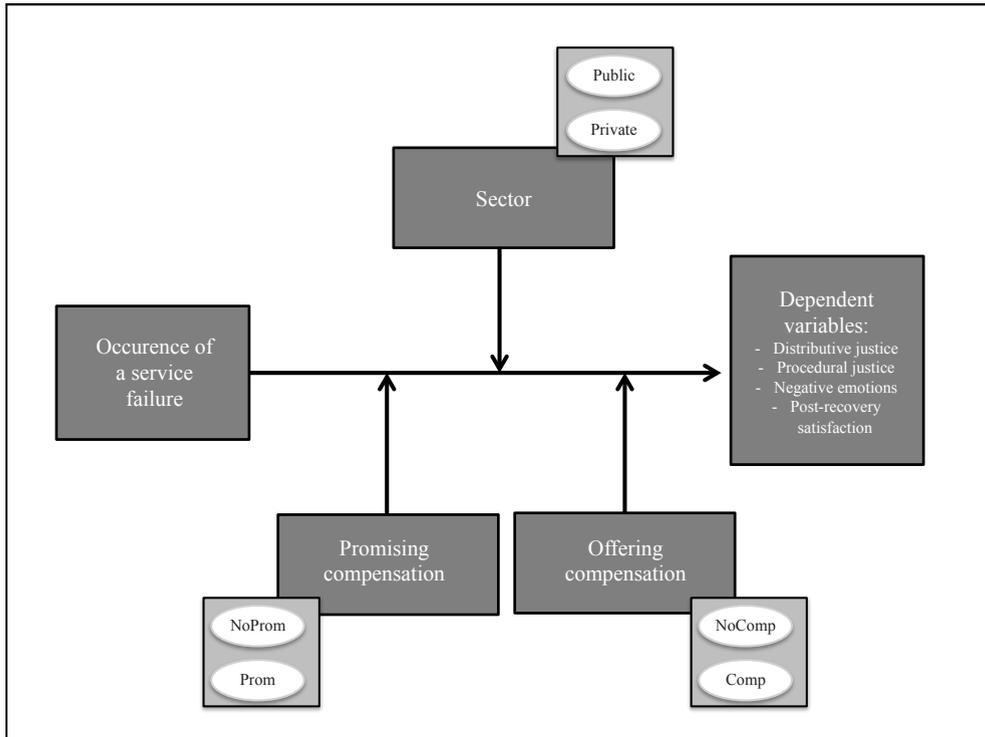
I expected that overall the effects would be similar in public and private settings for all four hypotheses. Citizens have other relationships with public services beyond those of a customer - a user, a voter, a recipient and a taxpayer (Milakovich, 2003). Also, as public organisations are often financed by taxpayers' money this will likely influence how customers evaluate public organisations offering compensation (Björlin Lidén and Edvardsson, 2003). However, in situations where customers pay directly for services, their money is directly related to the value they receive (Alford, 2002). Also, experiences in private settings might influence customers' public service expectations. Finally, increasing marketisation of public services, and the introduction of many private management and customer relation innovations, may have shifted customers' expectations to levels similar to those found in the private sector. Therefore, hypothesis 5 (H5) is: the effects as hypothesised in H1-2-3-4 are similar in public and private settings.

### **3.8. Overview of experiments**

I used two survey experiments involving students and a large and heterogeneous sample of US-citizens as proposed by Bouwman and Grimmelikhuijsen (2016). I first studied the hypotheses in an experiment with a student sample (for an overview of (dis)advantages of using students see Bouwman and Grimmelikhuijsen, 2016, p. 114). I recognise the limitations of a relatively homogeneous and small sample and to further validate results employed a follow-up replication study with a larger and more heterogeneous sample of US-citizens. The similarity of results allowed some theoretical generalisation. Both survey experiments used a 2 (sector: public, private) by 2 (compensation promised: no, yes) by 2 (compensation offered:

no, yes) between-subjects factorial design. Participants were randomly assigned to one of eight scenarios as graphically presented in Figure 2.

Figure 2. Graphical presentation of the (in)dependent variables



My aim was not to study types of service failures and recovery, other than the compensation aspect, which I needed to keep constant. Both experiments employed one service failure – service recovery configuration with a delay type of process failure that could be solved later (for a full description of scenarios, see Appendix II). The customer had ordered a product and was informed that it was ready for collection at a pick-up point. At the counter the customer found it was not there, but would be available the next day. Subsequently, I manipulated Promising Compensation by adding in the Prom-vignettes that the customer sees the promise: ‘We keep our promises, if not, you’ll get a gift voucher worth 5 dollar/euro’. I manipulated Offering Compensation by only offering the gift voucher in the Comp-vignettes. For monetary compensation I used a small and token compensation worth 5 dollar/euro that was given on the spot and proactively offered since the failure could be rectified the next day. It is of note that the way in which employees interact with customers in a recovery situation is

important and affects post-recovery satisfaction. However, the focus of my research was on the effects of compensation. I kept aspects related to interactional justice (such as offering an apology) constant in all scenarios. Therefore, interactional justice was not measured as a dependent variable.

### **3.9. Study 1: Student sample**

#### *Independent and dependent variables*

To manipulate Sector, I presented participants with a vignette within a municipality context (applied for a new driving license after expiration) or an Internet store (ordered a gift). I measured four dependent variables. An overview of the items and Cronbach's Alphas is given in Appendix III. Distributive justice was measured using an adapted Lii and Lee (2012) scale. For procedural justice and the proactive compensation offer I created a new three-item scale since current scales are based on situations where customers have to complain to receive compensation. Negative emotions were measured using the Mattila (2001) scale. For post-recovery satisfaction I used the scale applied by Huang and Lin (2011). As control variables, I asked participants to indicate the perceived severity of the failure (Mattila, 2001) and realism of the scenario (Magnini *et al.*, 2007). Finally, I asked participants three manipulation check questions verifying participant's understanding of the vignette. I tested and found that the data met all requirements to run factorial analyses of variance – ANOVAs (e.g. Field, 2013).

#### *Data collection, sample and validity check*

A total of 160 undergraduate students from a Dutch university participated for course credits. They visited the research laboratory and completed the on-line questionnaire in one-person cabins. Two students did not comply with the criterion of being Dutch, and were excluded because of possible national differences in reputations and relationships with public organisations. One participant failed two out of three manipulation checks and was excluded from the dataset. I included participants with one mistake, as the results were similar to exclusion. This resulted in 157 valid cases ( $M_{\text{age}} = 21.3$ ,  $SD = 2.0$ ; 43.9% female). For this initial test of the focal hypothesis, I took an investigative approach to the determination of sample size. An a priori power analysis using G\*power (Faul *et al.*, 2009) with a medium effect size ( $d = 0.25$ ) and power of 80%, indicated a sample size of 128 participants in total.

## Results

Control variables: a full factorial ANOVA on perceived severity of failure yielded no significant main or interaction effects. Also, age and gender had no significant effects on the dependent variables. An ANOVA on perceived realism of the scenario however yielded main Sector and Compensation effects. The internet store scenario ( $M_{internet\ store} = 4.79, SD = 1.52$ ) was perceived as significantly more realistic ( $F(1,149) = 22.53, p = .000$ ) as the municipality scenario ( $M_{municipality} = 3.62, SD = 1.79$ ). Also, not offering compensation ( $M_{no} = 4.88, SD = 1.61$ ) was perceived as significantly more realistic ( $F(1,149) = 30.70, p = .000$ ) as offering compensation ( $M_{yes} = 3.56, SD = 1.63$ ). This implies realism might explain the effects as well. I also ran ANCOVAs with realism as a covariate as a robustness check. These analyses resulted in similar results as presented below, only with main effects of Sector disappearing in ANCOVAs. But one of the assumptions of ANCOVAs is that the covariates (in this study for example ‘realism’) should not be dependent from the independent variables (see Miller and Chapman, 2001; Gerber and Green, 2012). Since this is the case in my study, I cannot use realism as a covariate and therefore I report the results of ANOVAs. I discuss this issue further in the general discussion.

General results: Table 5 summarises the results of the ANOVAs, including all main and interaction effects for all dependent variables. Below I discuss the effects directly related to hypotheses 1-5. Only H1 could be directly tested with these main and interaction effects alone, since the specific mean cell values have to be compared. For example, the double deviation condition Prom-NoComp has extreme low values and influences the overall Prom- and Comp-means tested in the main effects. To compare specific cell means and study the hypotheses 2-5, additional simple contrast analyses were used as part of the ANOVAs. For hypothesis 3 (NoProm-NoComp versus Prom-Comp) and for hypothesis 4b (Prom-NoComp versus NoProm-Comp) non-adjacent cell means were tested by separately calculating t-values based on the involved cell means, SDs and N per cell (cf. Lakens 2013). I used significance levels of 0.05 and confidence intervals of 95.0% throughout.

Table 5. Summary statistics of four ANOVAs (significant effects in bold)

|   | Distributive Justice   |                | Procedural Justice     |                | Negative Emotions      |                | Post-recovery Satisfaction |                |
|---|------------------------|----------------|------------------------|----------------|------------------------|----------------|----------------------------|----------------|
|   | <i>Wilks F (1,149)</i> | <i>p-value</i> | <i>Wilks F (1,149)</i> | <i>p-value</i> | <i>Wilks F (1,148)</i> | <i>p-value</i> | <i>Wilks F (1,148)</i>     | <i>p-value</i> |
| promising compensation                                  | 1.65                   | 0.201          | <b>21.36</b>           | <b>0.000</b>   | <b>9.85</b>            | <b>0.002</b>   | <b>5.40</b>                | <b>0.021</b>   |
| offering compensation                                   | <b>55.58</b>           | <b>0.000</b>   | <b>61.32</b>           | <b>0.000</b>   | <b>22.04</b>           | <b>0.000</b>   | <b>39.76</b>               | <b>0.000</b>   |
| sector  | 0.66                   | 0.417          | <b>7.35</b>            | <b>0.008</b>   | 0.81                   | 0.369          | <b>6.53</b>                | <b>0.012</b>   |
| sector * promising compensation                         | 0.38                   | 0.540          | 0.00                   | 0.977          | 0.55                   | 0.459          | 2.00                       | 0.160          |
| sector * offering compensation                          | 0.33                   | 0.567          | 0.30                   | 0.584          | 0.02                   | 0.887          | 2.85                       | 0.094          |
| promising compensation * offering compensation          | 1.15                   | 0.285          | <b>24.80</b>           | <b>0.000</b>   | <b>13.32</b>           | <b>0.000</b>   | 0.57                       | 0.452          |
| sector * promising compensation * offering compensation | 0.03                   | 0.873          | 0.00                   | 0.988          | 0.39                   | 0.536          | 1.75                       | 0.188          |

The effects of Offering Compensation (H1)

Results of the ANOVAs show that there are main Offering Compensation effects for all dependent variables. The pattern of means (Comp versus NoComp) moves as I hypothesised (see Table 6) concluding that H1 can be confirmed.

The effects of Promising Compensation (H2 and H3)

As Table 5 shows, Promising Compensation had a significant effect on procedural justice, negative emotions and post-recovery satisfaction. Of more interest was the effect of Promised Compensation, which was qualified by Offered Compensation (i.e. the interaction effect was significant) for procedural justice and negative emotions. To better understand this interaction, I used simple contrast analyses to test whether Offering Compensation without prior promises (NoProm-Comp) would lead to more positive evaluations when compared to where compensation is promised (Prom-Comp). The results showed that these two scenarios do not lead to significant differences in distributive justice ( $p = 0.881$ ;  $SE = .264$ ), procedural justice ( $p = 0.800$ ;  $SE = .292$ ), negative emotions ( $p = 0.716$ ;  $SE = .330$ ) and post-recovery satisfaction ( $p = 0.365$ ;  $SE = .225$ ). Hence, H2 cannot be confirmed.

Table 6. Observed means and standard deviation (between parentheses)

|  |               | n  | Distributive Justice | Procedural Justice | Negative Emotions | Post-recovery Satisfaction |
|--|---------------|----|----------------------|--------------------|-------------------|----------------------------|
| <b>Main effects</b>  |               |    |                      |                    |                   |                            |
| NoProm   |               | 77 | 3.93 (1.33)          | 4.75 (1.46)        | 3.64 (1.53)       | 3.02 (1.16)                |
| Prom   |               | 80 | 3.68 (1.39)          | 3.78 (1.77)        | 4.39 (1.69)       | 2.62 (1.13)                |
| NoComp   |               | 78 | 3.09 (1.10)          | 3.43 (1.54)        | 4.60 (1.67)       | 2.30 (0.97)                |
| Comp   |               | 79 | 4.50 (1.23)          | 5.08 (1.43)        | 3.47 (1.44)       | 3.31 (1.11)                |
| Public   |               | 77 | 3.72 (1.30)          | 3.97 (1.72)        | 4.14 (1.68)       | 2.61 (1.04)                |
| Private  |               | 80 | 3.87 (1.42)          | 4.54 (1.63)        | 3.92 (1.62)       | 3.01 (1.23)                |
| <b>Interaction effect Promising*Offering compensation</b>        |               |    |                      |                    |                   |                            |
|  | NoProm-NoComp | 38 | 3.32 (1.16)          | 4.46 (1.25)        | 3.77 (1.64)       | 2.57 (0.99)                |
|  | Prom-NoComp   | 40 | 2.88 (1.01)          | 2.46 (1.09)        | 5.38 (1.29)       | 2.05 (0.90)                |
|  | NoProm-Comp   | 39 | 4.52 (1.22)          | 5.04 (1.60)        | 3.53 (1.43)       | 3.44 (1.17)                |
|  | Prom-Comp     | 40 | 4.48 (1.25)          | 5.11 (1.25)        | 3.41 (1.46)       | 3.18 (1.05)                |
| <b>Interaction effect Sector*Promising*Offering compensation</b> |               |    |                      |                    |                   |                            |
| <b>Public</b>  | NoProm-NoComp | 18 | 3.22 (1.19)          | 4.22 (1.34)        | 3.88 (1.72)       | 2.83 (0.98)                |
|  | Prom-NoComp   | 20 | 2.93 (0.99)          | 2.23 (0.97)        | 5.45 (1.20)       | 1.90 (0.69)                |
|  | NoProm-Comp   | 19 | 4.34 (1.24)          | 4.70 (1.86)        | 3.82 (1.44)       | 3.78 (1.14)                |
|  | Prom-Comp     | 20 | 4.39 (1.15)          | 4.77 (1.22)        | 3.38 (1.64)       | 3.52 (1.16)                |
| <b>Private</b>   | NoProm-NoComp | 20 | 3.41 (1.16)          | 4.67 (1.15)        | 3.68 (1.61)       | 2.25 (0.93)                |
|  | Prom-NoComp   | 20 | 2.83 (1.05)          | 2.68 (1.17)        | 5.30 (1.40)       | 2.20 (1.06)                |
|  | NoProm-Comp   | 20 | 4.69 (1.21)          | 5.37 (1.27)        | 3.25 (1.40)       | 3.09 (1.11)                |
|  | Prom-Comp     | 20 | 4.56 (1.37)          | 5.45 (1.22)        | 3.45 (1.29)       | 2.85 (0.83)                |

To verify whether Prom-Comp and NoProm-Comp would result in more positive evaluations than NoProm-NoComp (H3), I used simple contrasts and t-tests. When combined, these reveal that H3 could be confirmed for post-recovery satisfaction, distributive and procedural justice, but not for negative emotions. Distributive justice for NoProm-Comp was significantly higher ( $p = 0.000$ ;  $SE = .268$ ) than for NoProm-NoComp. Also, Prom-Comp led to a significantly higher level than NoProm-NoComp ( $t(76) = 4.24$ ,  $p = 0.000$ ). Procedural justice for NoProm-Comp was significantly higher ( $p = 0.048$ ;  $SE = .295$ ) than NoProm-NoComp. Also, procedural justice for Prom-Comp was significantly higher than NoProm-NoComp ( $t(76) = 2.30$ ,  $p = 0.025$ ). However, there were no differences on negative emotions between NoProm-NoComp and NoProm-Comp ( $p = 0.467$ ;  $SE = .337$ ), nor with Prom-Comp ( $t(76) = -1.03$ ,  $p = 0.309$ ). Finally, both participants in NoProm-Comp ( $p = 0.000$ ;  $SE = .230$ ) and in Prom-Comp ( $t(76) = 2.64$ ,  $p = 0.010$ ) scenarios were significantly more satisfied than in NoProm-NoComp scenarios.

#### The effects of double deviation (H4)

I used simple contrast analyses to test whether a double deviation, i.e. Prom-NoComp, would be significantly more negative than NoProm-NoComp. H4a could be confirmed for procedural justice ( $p = 0.000$ ;  $SE = .294$ ), negative emotions ( $p = 0.000$ ;  $SE = .335$ ; for cell means, see Table 6) and post-recovery satisfaction ( $p = 0.032$ ;  $SE = .229$ ), but not for distributive justice ( $p = 0.099$ ;  $SE = .266$ ).

In hypothesis H4b, I stated that for all four dependent variables Prom-NoComp would lead to significant more negative evaluations than for NoProm-Comp and Prom-Comp. I used simple contrast analyses to compare Prom-NoComp with Prom-Comp, and separate t-tests to compare Prom-NoComp with NoProm-Comp. The results show that this hypothesis can be confirmed for all dependent variables. Prom-NoComp does lead to a significant lower distributive justice than Prom-Comp ( $p = 0.000$ ;  $SE = .263$ ) and NoProm-Comp ( $t(77) = -6.52$ ,  $p = 0.000$ ). Procedural justice in the Prom-NoComp was significantly lower compared to both Prom-Comp ( $p = 0.000$ ;  $SE = .290$ ) and NoProm-Comp ( $t(77) = -8.39$ ,  $p = 0.000$ ). Negative emotions for Prom-NoComp were significantly higher than for Prom-Comp ( $p = 0.000$ ;  $SE = .328$ ) and NoProm-Comp ( $t(77) = 6.04$ ,  $p = 0.309$ ). Finally, post-recovery satisfaction for Prom-NoComp was significantly lower than for Prom-Comp ( $p = 0.000$ ;  $SE = .224$ ) and NoProm-Comp ( $t(77) = -5.93$ ,  $p = 0.010$ ).

#### Differences between two sectors (H5)

To study whether Sector made a difference in the hypothesised effects of promising and offering compensation, I looked at the main and interaction effects involving the sector. Sector did not moderate the dependent variables by interaction effects with promising or offering compensation, suggesting H1-4 are true for both public and private settings. I did find significant main Sector effects (see Table 5) for procedural justice ( $M_{\text{public}} = 3.97$ ,  $SD = 1.72$ ;  $M_{\text{private}} = 4.54$ ,  $SD = 1.63$ ) and post-recovery satisfaction ( $M_{\text{public}} = 2.61$ ,  $SD = 1.04$ ;  $M_{\text{private}} = 3.01$ ,  $SD = 1.23$ ). However, I must be careful interpreting these effects given that Sector also influenced our control variable. I will elaborate on this in the General Discussion.

### 3.10. Study 2: Online US panel

#### *Independent and dependent variables*

As recommended by Bouwman and Grimmelikhuijsen (2016) I replicated study 1 with a larger, more heterogeneous sample employing the same stimuli, but with one change. Because

of differences in the process of acquiring driving licenses between The Netherlands and USA, for the public setting I selected a governmental organisation that issues visa (see Appendix II). Promising and offering compensation were manipulated similar to study 1. Also, the dependent variables were identical to those in study 1 (for items and reliability measures, see Appendix III).

#### *Data collection, sample and validity check*

Based on the findings in Study 1, I anticipated a small effect size ( $d = .10$ ). Power analyses using G\*power (Faul *et al.*, 2009) for the 8-condition design ( $\alpha = .05$  and power of 80%), suggested to at least have 787 participants. Because of potential dropouts of participants, I recruited a heterogeneous set of participants ( $N = 1055$ ) via Amazon's Mechanical Turk (MTurk) online panel. This is a suitable sampling frame for public experiments according to Jilke *et al.*, (2016) and Stritch *et al.*, (2017). The MTurk population (MTurkers) is not a random sample of the US-population and hence not statistical representative. However, according to Bouwman and Grimmelikhuijsen (2016) experiments do not necessarily have to rely on random samples. The population of MTurkers is compared with student samples and standard internet panels very diverse in terms of demographic characteristics (Buhrmester *et al.*, 2011). Also, considerable research has shown that MTurk research replicated in surveys, experimental studies and behavioural laboratory research found few substantial differences (Jilke *et al.*, 2016). Another criticism is that MTurk workers are paid so little; one may wonder whether they take the experiment seriously (Paolacci *et al.*, 2010). But Hauser and Schwarz (2016) showed in three online studies that MTurkers are more attentive to instructions than student pools.

To tackle the possible effects of this attentiveness problem, to ensure a high reliability of the dataset and increase statistical power, I employed an identical procedure to that used by Jilke *et al.*, (2016). First, I excluded 43 participants based on wrong answers of an instructional manipulation check question. Second, 17 respondents with two mistakes out of three manipulation check questions on the sector, promising and offering compensation were excluded. I included participants with one mistake, as the results were similar to exclusion. The highest and lowest 1% percentile in terms of total survey completion time ( $N=16$ ) and 33 respondents with an overlapping IP-address were excluded. Finally, 9 non-US citizens were excluded from the dataset. This led eventually to a sample of 937 valid cases ( $M_{\text{age}} = 38.3$ ,  $SD = 12.35$ ; 47.8% female; US nationality), ensuring sufficient power.

## Results

Control variables: a full factorial ANOVA on perceived severity of the failure yielded a main Sector effect ( $F(1,928) = 6.355, p = .012; M_{visa} = 5.36, SD = 1.46; M_{internet\ store} = 5.60, SD = 1.29$ ). An ANOVA on perceived realism yielded a main Sector ( $F(1,929) = 13.205, p = .000; M_{visa} = 5.61, SD = 1.47; M_{internet\ store} = 5.93, SD = 1.14$ ) and a main Compensation effect ( $F(1,929) = 26.525, p = .000; M_{yes} = 5.55, SD = 1.47; M_{no} = 5.99, SD = 1.12$ ). Both sector and realism can have a significant effect on the dependent variables. Also for this study I ran ANCOVAs as robustness checks showing only minor differences in Sector effects. The limitations are discussed in the general discussion. As in study 1, age and gender did not have significant effects on the dependent variables.

General results: Table 7 summarises the results of the ANOVAs including all main and interaction effects for all dependent variables. As in study 1, I discuss these effects directly related to the hypotheses 1-5. Similar to the first study, I used additional simple contrast analyses and separate t-tests to compare means of different scenarios. For all our analyses, I used significance levels of 0.05 and confidence intervals of 95.0%.

Table 7. Summary statistics of four ANOVAs (significant effects in bold)

|   | Distributive Justice |              | Procedural Justice |              | Negative Emotions |              | Post-recovery Satisfaction |              |
|---|----------------------|--------------|--------------------|--------------|-------------------|--------------|----------------------------|--------------|
|   | Wilks F (1,929)      | p-value      | Wilks F (1,929)    | p-value      | Wilks F (1,929)   | p-value      | Wilks F (1,929)            | p-value      |
| promising compensation                                  | 1.88                 | 0.171        | <b>16.72</b>       | <b>0.000</b> | 0.09              | 0.767        | 0.58                       | 0.446        |
| offering compensation                                   | <b>407.37</b>        | <b>0.000</b> | <b>655.85</b>      | <b>0.000</b> | <b>122.55</b>     | <b>0.000</b> | <b>299.129</b>             | <b>0.000</b> |
| sector  | 0.74                 | 0.391        | 1.34               | 0.237        | <b>5.00</b>       | <b>0.026</b> | 1.44                       | 0.230        |
| sector * promising compensation                         | 1.89                 | 0.196        | <b>6.15</b>        | <b>0.013</b> | 0.19              | 0.663        | 0.21                       | 0.645        |
| sector * offering compensation                          | 0.89                 | 0.345        | 0.31               | 0.578        | 0.67              | 0.412        | 0.85                       | 0.357        |
| promising compensation * offering compensation          | <b>6.66</b>          | <b>0.010</b> | <b>60.12</b>       | <b>0.000</b> | 2.66              | 0.103        | 2.41                       | 0.121        |
| sector * promising compensation * offering compensation | 2.39                 | 0.123        | <b>5.09</b>        | <b>0.024</b> | 0.05              | 0.821        | 0.05                       | 0.816        |

The effects of Offering Compensation (H1)

ANOVAs for the dependent variables yielded significant main effects on all dependent variables (see Table 7). The pattern of means was as expected by the hypothesis (see Table 8), concluding that H1 can be confirmed.

The effects of Promising Compensation (H2 and H3)

As Table 7 shows, there is a main effect of Promising Compensation on procedural justice. However, I also see that this main effect is qualified by an interaction effect of promising and offering compensation on distributive and procedural justice. Simple contrast analyses to test whether NoProm-Comp would lead to more positive evaluations than Prom-Comp (H2) showed that these two scenarios do not lead to significant differences in distributive justice ( $p = 0.393$ ;  $SE = .134$ ), negative emotions ( $p = 0.173$ ;  $SE = .126$ ) and post-recovery satisfaction ( $p = 0.576$ ;  $SE = .113$ ). However, there is a significant difference for procedural justice ( $p = 0.010$ ;  $SE = .125$ ). Hence, H2 can only be confirmed for procedural justice.

Table 8. Observed means and standard deviation (between parentheses)

|  | n   | Distributive Justice | Procedural Justice | Negative Emotions | Post-recovery Satisfaction |
|--|-----|----------------------|--------------------|-------------------|----------------------------|
| <b>Main effects</b>  |     |                      |                    |                   |                            |
| NoProm   | 470 | 3.20 (1.69)          | 3.77 (1.66)        | 5.55 (1.43)       | 2.43 (1.42)                |
| Prom   | 467 | 3.07 (1.80)          | 3.42 (1.92)        | 5.52 (1.48)       | 2.37 (1.38)                |
| NoComp   | 469 | 2.18 (1.27)          | 2.47 (1.38)        | 6.03 (1.17)       | 1.71 (0.94)                |
| Comp   | 468 | 4.09 (1.62)          | 4.72 (1.43)        | 5.03 (1.54)       | 3.09 (1.45)                |
| Public   | 476 | 3.21 (1.78)          | 3.68 (1.83)        | 5.42 (1.50)       | 2.47 (1.44)                |
| Private  | 461 | 3.06 (1.70)          | 3.50 (1.78)        | 5.65 (1.39)       | 2.33 (1.37)                |
| <b>Interaction effect Promising*Offering compensation</b>        |     |                      |                    |                   |                            |
| NoProm-NoComp  | 236 | 2.36 (1.24)          | 2.98 (1.41)        | 5.97 (1.21)       | 1.80 (1.00)                |
| Prom-NoComp  | 233 | 1.99 (1.29)          | 1.95 (1.14)        | 6.09 (1.13)       | 1.62 (0.86)                |
| NoProm-Comp  | 234 | 4.04 (1.67)          | 4.56 (1.52)        | 5.12 (1.52)       | 3.06 (1.50)                |
| Prom-Comp  | 234 | 4.15 (1.57)          | 4.88 (1.32)        | 4.95 (1.56)       | 3.12 (1.40)                |
| <b>Interaction effect Sector*Promising*Offering compensation</b> |     |                      |                    |                   |                            |
| <b>Public</b>  |     |                      |                    |                   |                            |
| NoProm-NoComp  | 115 | 2.59 (1.28)          | 3.27 (1.34)        | 5.80 (1.31)       | 1.90 (1.03)                |
| Prom-NoComp  | 116 | 1.94 (1.30)          | 1.81 (1.01)        | 5.98 (1.16)       | 1.69 (0.96)                |
| NoProm-Comp  | 119 | 4.03 (1.80)          | 4.60 (1.61)        | 5.05 (1.60)       | 3.10 (1.59)                |
| Prom-Comp  | 126 | 4.16 (1.62)          | 4.90 (1.39)        | 4.90 (1.60)       | 3.11 (1.41)                |
| <b>Private</b>   |     |                      |                    |                   |                            |
| NoProm-NoComp  | 121 | 2.14 (1.15)          | 2.70 (1.41)        | 6.13 (1.09)       | 1.71 (0.98)                |
| Prom-NoComp  | 117 | 2.05 (1.28)          | 2.08 (1.24)        | 6.19 (1.08)       | 1.54 (0.76)                |
| NoProm-Comp  | 115 | 4.05 (1.53)          | 4.52 (1.42)        | 5.20 (1.42)       | 3.02 (1.41)                |
| Prom-Comp  | 108 | 4.15 (1.53)          | 4.86 (1.24)        | 5.00 (1.53)       | 3.14 (1.39)                |

In H3 I hypothesised that both compensation scenarios (Prom-Comp and NoProm-Comp) would lead to more positive evaluations than NoProm-NoComp. As in study 1, additional simple contrast analysis and specific t-tests were used to compare means of these three scenarios involved. These confirm H3 for all dependent variables. Distributive justice for NoProm-Comp is significantly higher ( $p = 0.000$ ;  $SE = .134$ ) than for NoProm-NoComp. Also, Prom-Comp leads to a significantly higher level than NoProm-NoComp ( $t(468) = 13.72$ ,  $p = 0.000$ ). Procedural justice for the NoProm-Comp scenario is significantly higher ( $p = 0.000$ ;  $SE = .124$ ) than NoProm-NoComp. Also, procedural justice for the Prom-Comp scenario is significantly higher than NoProm-NoComp ( $t(468) = 15.08$ ,  $p = 0.000$ ). For negative emotions, NoProm-Comp leads to significantly less negative emotions ( $p = 0.000$ ;  $SE = .126$ ) than NoProm-NoComp. This is also the case for Prom-Comp ( $t(468) = -7.92$ ,  $p = 0.000$ ). Finally, both NoProm-Comp ( $p = 0.000$ ;  $SE = .113$ ) and Prom-Comp ( $t(468) = 11.77$ ,  $p = 0.000$ ) lead to significant higher levels of post-recovery satisfaction than NoProm-NoComp.

The effects of double deviation (H4)

Testing whether Prom-NoComp (double deviation) would lead to significantly more negative evaluations than NoProm-NoComp (H4a) analysing the interaction effects with simple contrast analyses reveals that this could be confirmed for distributive justice ( $p = 0.005$ ;  $SE = .134$ ) and procedural justice ( $p = 0.000$ ;  $SE = .124$ ), but not for negative emotions ( $p = 0.345$ ;  $SE = .126$ ) and post-recovery satisfaction ( $p = 0.102$ ;  $SE = .113$ ).

Simple contrast analysis and specific t-tests revealed that hypothesis H4b can be confirmed for all dependent variables. Prom-NoComp leads to a significantly lower distributive justice than Prom-Comp ( $p = 0.000$ ;  $SE = .134$ ) and NoProm-Comp ( $t(465) = -14.84$ ,  $p = 0.000$ ). Procedural justice for Prom-NoComp is significantly lower than Prom-Comp ( $p = 0.000$ ;  $SE = .125$ ) and NoProm-Comp ( $t(465) = -20.98$ ,  $p = 0.000$ ). Negative emotions for Prom-NoComp are significantly higher than for Prom-Comp ( $p = 0.000$ ;  $SE = .127$ ) and NoProm-Comp ( $t(465) = 7.82$ ,  $p = 0.000$ ). Finally, post-recovery satisfaction for Prom-NoComp is significantly lower than for Prom-Comp ( $p = 0.000$ ;  $SE = .113$ ) and NoProm-Comp ( $t(465) = -12.72$ ,  $p = 0.000$ ).

### Differences between two sectors (H5)

To study the Sector effects I looked at the main and all interaction effects involving Sector (see Table 7). As Table 6 shows, there is a main Sector effect on negative emotions. The private scenarios ( $M = 5.65$ ;  $SD = 1.39$ ) lead to significantly ( $F(1,929) = 5.00$ ,  $p = 0.026$ ) more negative emotions than the public organisation ( $M = 5.42$ ;  $SD = 1.50$ ). Overall, the data suggest that Sector did not seem to affect customers' perceived distributive justice, procedural justice or post-recovery satisfaction, thus confirming H5 for these dependent variables. However, I also see that there is an interaction effect ( $F(1,929) = 6.15$ ,  $p = .013$ ) of Sector and Promising Compensation on procedural justice. A simple contrast analysis shows that not promising compensation leads to a significantly ( $p = 0.010$ ;  $SE = .124$ ) higher level for the public ( $M = 3.94$ ;  $SE = 0.088$ ) than for the private scenario ( $M = 3.61$ ;  $SE = 0.088$ ). For the scenarios promising compensation does not lead to significant differences ( $p = 0.360$ ;  $SE = .125$ ). Finally, for procedural justice there is an interaction effect between Sector, Promising and Offering compensation ( $F(1,929) = 5.09$ ,  $p = .024$ ). A close inspection of the pattern of means (see Table 7), shows that within the private setting, there is only an effect of offering compensation: People perceive lower procedural justice when there is no compensation ( $M = 2.70$  and  $2.08$  respectively) than when there is compensation ( $M = 4.52$  and  $4.86$  respectively). However, this pattern is different within the public domain. Promising compensation, but not offering it (Prom-NoComp) leads to much lower perceived procedural justice ( $M = 1.81$ ) compared to NoProm-NoComp ( $M = 3.27$ ). There is no difference between NoProm-Comp ( $M = 4.60$ ) and Prom-Comp ( $M = 4.90$ ). Compared to Study 1, these results seem to suggest that Sector might moderate some of the effects offering and promising compensation has. I do need to emphasise though, that similar to Study 1, the conclusions of these findings need to be taken with caution given the effect Sector had on other control variable.

### 3.11. Summary of two studies

The results of both studies are summarised in Table 8 showing that hypotheses 1, 2, 3, 4b and 5 have many similarities. For H4a the results of both studies are mixed.

Table 9. Results of two studies

|   | Study 1  | Study 2   |
|---|--|---|
| <i>H1. Comp leads to more positive evaluations than NoComp</i>                            | Confirmed: all dependent variables.  | Confirmed: all dependent variables.   |
| <i>H2. Prom-Comp leads to more negative evaluations than NoProm-Comp</i>                  | Not confirmed: all dependent variables.  | Confirmed: procedural justice.<br>Not confirmed: distributive justice, negative emotions, post-recovery satisfaction.                                 |
| <i>H3. Prom-Comp and NoProm-Comp lead to more positive evaluations than NoProm-Comp</i>   | Confirmed: distributive and procedural justice, post-recovery satisfaction.<br>Not confirmed: negative emotions.         | Confirmed: all dependent variables.   |
| <i>H4a. Prom-NoComp leads to less positive evaluations than NoProm-NoComp</i>             | Confirmed: procedural justice, negative emotions and post-recovery satisfaction.<br>Not confirmed: distributive justice. | Confirmed: distributive and procedural justice<br>Not confirmed: negative emotions and post-recovery satisfaction.                                    |
| <i>H4b. Prom-NoComp leads to less positive evaluations than Prom-Comp and NoProm-Comp</i> | Confirmed: all dependent variables.  | Confirmed: all dependent variables.   |
| <i>H5. Results of H1-H4 are similar for public and private settings</i>                   | Confirmed (but with caution): all dependent variables.   | Confirmed (but with caution): all dependent variables, except for interaction effects with Promising and Offering Compensation on procedural justice. |

### 3.12. Discussion

Two experiments amongst Dutch students (study 1) and US citizens (study 2) researched the effects of promising and offering a small monetary compensation after a service failure in public and private settings. Eight scenarios were used to determine the effects of these independent variables on customers' evaluations of distributive and procedural justice, negative emotions and post-recovery satisfaction. The results let me conclude that offering compensation after a service failure significantly improved customers' evaluations. Promising compensation explicitly before a service failure had no effect on these evaluations. When compensation was promised, but not offered, this led to significantly more negative evaluations than when it was offered. Finally, the sector (public vs. private) did not seem to moderate the effects of promising and offering compensation on customers' evaluations (with one exception in Study 2).

This research adds to public management literature. It is the first study to apply justice theory (Adams, 1965) in public service recovery settings and proves itself relevant. The results show monetary compensation is a powerful instrument to restore customers' perceived loss and contribute to the restoration of satisfaction after a service failure; an effect already found in private settings (e.g. Wirtz and Mattila, 2004) and now confirmed in public settings. However, from an ethical and legal perspective, offering individual customers compensation may not always be desirable or possible. One reason is that such compensation has to be paid from general tax revenues. This benefits individual citizens and could lead to inequalities and

stimulate a claim culture. Also, customers as taxpayers could object to public organisations using tax money to compensate customers.

Second, based on expectancy disconfirmation theory (Oliver, 1993) I expected that explicitly promising compensation prior to the failure would lead to more negative customers' evaluations compared to when it has not been promised. This is based on the effect that expectations increase, while experiences stay the same. My results show that this is not the case. From a service recovery perspective, promising compensation has no effects on customers' evaluations. However, expectancy disconfirmation theory is clearly supported in the case of a double deviation where the promised compensation is not offered. When expectations of receiving compensation are set and, subsequently, not fulfilled, this leads to very negative evaluations. From a managerial perspective, the question is: what sense does it make to promise compensation explicitly? Where no second failures are made, it does not lead to better evaluations. In the case of a double deviation, it leads to extreme negative evaluations. Finally, I expected that, in the direct-exchange scenarios where customers directly pay for services, there would be no difference in customers' evaluations between the public and private settings. Although I found some main and interaction effects involving sector, I can conclude that the sector has no mediating effects on customers' evaluations.

My study also has some limitations, which could lead to new lines of future research. First, while I tried to develop realistic scenarios with similar failures with a similar severity, participants of both studies perceived the private scenarios more realistic than the public scenarios; probably because offering service guarantees and monetary compensation are not yet common public practices. This scenario could be re-examined when service guarantees and monetary compensation do become more common to see whether the difference in perceived realism will disappear. Also, because I did not anticipate this effect on perceived realism, I did not disentangle whether the impact of positive and normative expectations (cf. James, 2011) on customers' evaluations are related to realism. For now, I need to interpret the significant effects of Sector with caution, because they could be partly explained in terms of perceived realism rather than Sector itself. I therefore also cannot draw strong conclusions regarding to H5 (difference in effects between private vs. public). Second, I intentionally kept the interaction with the employee (interactional justice) constant in order to focus on the effects of compensation and the payout process. In follow-up studies the effects of employees behaviour on the effects of compensation could be researched. Third, I simulated using one

type and level of severity of service failure and one type and size of monetary compensation. Future research could investigate whether different types of failures and compensation, different severity levels of failures and different sizes of compensation influence customers' evaluations. It could also look for the specific effects of promising compensation on customers' expectations (e.g. could size of compensation moderate the effect) to understand better the impact of expectations. To expand the knowledge on the effects of compensation future research could study public services (other than governmental visa organisations and municipalities issuing driving licenses) with a broad spectrum of all public customer segments. Finally, research could be conducted in non-direct exchange situations where public services are paid with taxpayers' money. Despite these limitations, my research has given early insights into the effects of promising and offering monetary compensation in public service recovery situations.