Available online at www.sciencedirect.com
SCIENCE DIRECT®
Organizational Behavior and Human Decision Processes 97 (2005) 1–17
ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
www.elsevier.com/locate/obhdp

Restoring trustworthiness after adverse events: The signaling effects of voluntary "Hostage Posting" on trust ☆

Kazuya Nakayachi [a,*], Motoki Watabe [b]

[a] *Tezukayama University, Nara, Japan*
[b] *Kyoto University, Kyoto, Japan*

Received 8 July 2003
Available online 17 March 2005

Abstract

The present research investigates the effects of voluntary hostage posting by organizations—with the provisions of monitoring and self-sanctions—in order to restore public trust after adverse events. The results of the first two studies demonstrate that voluntary hostage posting raised participants' perceptions of the trustworthiness of organizations that had caused incidents, whereas imposed or involuntary hostage posting did not result in more positive evaluations. The third study revealed that voluntary posting affects not only the perception of trustworthiness but also respondents' behavior when their interests are at stake. These findings are consistent with a study by Slovic (1993), which suggested that the best way to increase public trust in a nuclear power plant was to delegate authority to shut down the plant to an outside monitor. Implications of these results for a theory of trust and management policy for restoring trust are discussed.

Keywords: Trust; Hostage Posting; Risk perception; Signaling effect; Trust game

Introduction

Recent social science literature has paid a great deal of attention to the topic of trust. In the field of risk management, for example, public trust has been emphasized because it has much influence on the response to technologies and human activities, as well as on policies for managing potential risk (Cvetkovich & Lofstedt, 1999; Cvetkovich, Siegrist, Murray, & Tragesser, 2002; Earle & Cvetkovich, 1995; Siegrist, 2000; Siegrist & Cvetkovich, 2000; Slovic, 1993, 1999). Among factors which damage the public trust, the most common are incidents caused by a company or government. This article addresses what companies and governments should do in order to avoid the loss of public trust after adverse events for which they are responsible. The literature shows that trust is easy to destroy and difficult to build (Kraus, Malmfors, & Slovic, 1992; Koren & Klein, 1991; Slovic, 1993; Slovic, Flynn, Johnson, & Mertz, 1993). Given that trust can be destroyed so easily, it is important to investigate what is needed to restore trust after an adverse event. Policy makers in companies and governmental agencies often want a prescription from social scientists for how to resolve a crisis of trust. Not many researchers, however, approach this problem directly (Schweizter, Hershey, & Bradlow, 2003). An empirical approach to recovery of trust from a dynamic viewpoint could be expected not only to produce suggestions for

☆ The authors would like to thank George Cvetkovich, Takashi Kusumi, the editor David Harrison and three anonymous reviewers for their helpful comments; we would also like to thank Mari Uchiumi, Kazumi Renge, Yoshiyuki Ueda and Hanae Yamada for their help with data collection as well as Kate Marrone for her careful proofread of the manuscript. This research was supported by Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science including 21 COE program for Kyoto University Psychology Union.
* Corresponding author. Fax: +81 742 41 4895.
E-mail address: nakayachi@tezukayama-u.ac.jp (K. Nakayachi).

0749-5978/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.obhdp.2005.02.001

management concerning trust in practice, but also to provide information for theoretical development in trust research.

In this article, we examine the effects of "Hostage Posting" on restoring perceived trustworthiness of the posters. Usually, there is some uncertainty regarding trustworthiness in social exchanges involving economic and social resources. Hostage posting, which is a commonly used term in Economics, refers to a self-sanctioning system in an uncertain situation. In research concerning social exchange, hostage posting has drawn considerable attention as a commitment device to reduce uncertainty and to resolve social conflict. In the two-person hostage game, for example, each of two actors decides on two moves. In the first stage, both actors decide independently whether to post a hostage, that is to deposit a considerable sum of money which would be forfeited in case they choose to defect on the subsequent move. In the second stage, each actor is informed of the other's first move (the decision regarding hostage posting) and decides independently whether to choose cooperation or defection. To post a hostage at the first stage changes the payoff structure and diminishes the incentive for actors to defect at the second stage. For example, let us say actor A posts an amount of money at the first stage which exceeds that which will be gained if actor A chooses defection and actor B chooses cooperation in the second stage. By doing so, the incentive for actor A to choose defection disappears. This situation would make actor B expect that actor A would not cheat because no incentive for defection remains. Thus, the expectation that the hostage poster will not cheat in this game situation would increase. Theoretical and empirical studies of social exchange situations have supported this trust-increasing effect of hostage posting (Gautschi, 1999; Keren & Raub, 1993; Mlicki, 1996; Raub & Keren, 1993; Williamson, 1983).

In our current research we apply hostage posting to the context of endangered trust after adverse events and examine theoretically and empirically its effect on the restoration of the perceived trustworthiness of the hostage posters. In the field of risk management as well as in organizational behavior, there has been no previous research investigating the trust-increasing effects of hostage posting, using its terms explicitly. Some research, however, could be interpreted from the viewpoint of hostage posting. For example, the results of a questionnaire-based study by Slovic (1993) revealed that the only event to have a substantial impact on increasing the estimation of trustworthiness for a nuclear power plant was that "an advisory board of local citizens and environmentalists be established to monitor the plant and be given legal authority to shut the plant down if they believe it to be unsafe." Such a strong delegation of authority to the local public can be interpreted as hostage posting, because this delegation increases the incentive for the plant to behave honestly and cooperatively towards the public. Nakayachi and Ohnuma (2003) also found that public perception of trustworthiness increases when the agency in charge of environmental management promotes the release of information and accepts monitoring and sanctions in case of deception. This acceptance diminishes the incentive to lessen circumspect management or to deceive the public. Therefore, their findings can be interpreted as an example of the effect of voluntary hostage posting on perceived trustworthiness. As these studies showed, one representative form of hostage posting by the organizations responsible is the acceptance of monitoring and sanctions by the organization themselves. On the other hand, the introduction of imposed monitoring and sanction rules does not always improve the public trust. For example, the Japanese government greatly strengthened the monitoring system for shipping beef after a BSE cow was found in Japan. With such tight security, a company trying to ship BSE meat would have been exposed and received a deathblow from the government and the market. Even with such potential sanctions, however, the public trust in beef producers was not recovered for a long time and consumers continued to not eat beef. This raises the question, what forms of hostage posting are most effective for the restoration of trustworthiness of the hostage posters?

Signaling effect of voluntary Hostage Posting

Our research shows that there seems to be another reason why hostage posting increases trust in a game situation, besides a change in the payoff structures. Posting a hostage means *the posters themselves* are creating a situation where they have little incentive to be deceptive because deceptive behavior would cause the forfeit of the hostage. Therefore, hostage posting not only changes incentive structures but also implies that the hostage posters are trustworthy because they are willing to bear the cost of the hostage and restrict their own self-interest in order to resolve social conflict. This function of hostage posting can be interpreted as a signaling effect. The significant aspect is that it is only when the trustees post a hostage voluntarily that hostage posting signals the posters' positive intentions. The mere acceptance of posting a hostage in response to trusters' demands does not affect the perceived trustworthiness of the trustees even though, just as in the case of voluntary hostage posting, it changes the payoff structure. In other words, perceived trustworthiness of the trustees would not be affected by the passive acceptance of posting a hostage. Posting a voluntary hostage also indicates that the posters are competent, which is one of the important components of trustworthiness. Trusters would infer that the trustees are posting a hostage voluntarily because they are confident that they can make good management decisions and that their hostage will not be forfeited.

As described above, one of our assertions in this article is that, after an incident endangering trust, the voluntariness in hostage posting determines whether or not hostage posting restores the perceived trustworthiness of the posters. In a classic analysis of conflict resolution, Shelling (1960) described voluntariness as one of the essential features explaining why hostage posting works as a commitment device to induce cooperation in a conflict situation. However, a detailed model of how the voluntariness of hostage posting influences trust has not yet been developed. Mlicki (1996) referred to the signaling effects of voluntary hostage posting, but he has not produced an empirical study regarding the signaling effects. In this paper, we investigate empirically the effects of voluntary hostage posting on perceived trustworthiness in the context of the endangered trust of a company after an adverse event.

The central thesis of this research, that voluntariness in posting a hostage is the determinant factor necessary to restore perceived trustworthiness, might seem to be obvious. However, models of hostage posting have not yet been directly applied in the empirical research of trust recovery. We are therefore trying to expand the model by adding the factor of voluntariness. Voluntariness in cooperative behavior cannot be explained only in the context of the payoff structure, as in economic theories, because the payoff structure is identical whether it is introduced voluntarily by the trustees or it is imposed by the trusters. However, studies based on psychology can examine the effects of voluntariness in hostage posting.

Furthermore, we do not presume that the effects of voluntary hostage posting concerning public trust are well understood and utilized in practice. Companies often wait passively for a response by overseeing authorities after an incident which endangers public trust. Our assertion, if it is proved to be true, suggests that companies should voluntarily introduce a monitoring and sanction rule before it is demanded by stakeholders. If the cost of providing a monitoring system is equivalent, the restoration of trustworthiness due to a voluntary offer of a hostage is far preferable for a company or a government than simply agreeing to provide it, which does not increase the perception of trustworthiness.

Factors of trustworthiness

A number of researchers have pointed out that the concept of trust is so diverse and different among disciplines that research in trust has been hampered (Barber, 1983; Cvetkovich & Winter, in press; Fischhoff, 1999; Mayer & Davis, 1999; Rousseau, Sitkin, Burt, & Camerer, 1998; Yamagishi, 1998). However, current definitions of trust emphasize the "willingness to be vulnerable to the actions of another party" proposed by Mayer, Davis, and Schoorman (1995). They noted two factors which affect trust, propensity of the truster and the trustworthiness of the trustee. They defined trustworthiness as an attribute of the trustee which determines the degree of trust. Adopting their definition, this research examines the effects of hostage posting on the perceived trustworthiness of a company responsible for an incident (trustee) by the public (truster).

Corresponding to the diverse definitions of trust, a wide range of factors of trustworthiness has been proposed. In the field of organizational behavior, Mayer et al. (1995) extracted three major characteristics of trustworthiness from the literature—ability, benevolence, and integrity. Mayer and Davis (1999) confirmed the constructive validity of this model in their field quasi-experiment. However, in other research areas, for example in risk management studies, the factors or dimensions remain controversial.

While the ability factor is fairly common in the literature, other factors seem to vary, for example, openness (Covello, 1992), fairness (Cummings & Bromiley, 1996), integrity (Slovic, 1993), and care (March & Olson, 1989). Other researchers have divided these factors into sub-components such as "openness and honesty" and "concern and care" (Peters, Covello, & McCallum, 1997), or a mixture of affective components (Metley, 1999). While these factors have varied, all of them seem to be connected to the aspect of trustees having a positive motivation, and of their not intending to deceive the trusters. This aspect seems to correspond to the concepts of benevolence and integrity in Mayer et al. (1995). Thus, we tentatively note two factors that denote trustworthiness, ability and motivation. However, the purpose of this article is to examine the effects of hostage posting on the public perception of trustworthiness of organizations responsible rather to propose a conclusive and general trustworthiness construct. Therefore, we changed the items related to the motivational factor between two questionnaire-based studies to see whether the results are general and replicable. In the first study, we used five items for the motivational factor that mainly reflected the integrity concept (trustworthiness, honesty, conscience, and regret), and one benevolence item (care about consumers). In the second study, we expanded the items for the motivational factor to 10. Half of them reflected the integrity concept (trustworthiness, honesty, keeping promise, responsibility, and reliability) and the other half reflected the benevolence concept (concern about interests, safety and thoughts of consumers; consideration of consumers' viewpoint; and efforts towards consumers' gratification).

According to the discussion above, we set up the first hypothesis regarding the effects of voluntary hostage posting.

Hypothesis 1. Voluntary hostage posting will raise the perceived trustworthiness of the posters compared to imposed hostage posting.

In addition to the perception of trustworthiness, we measured participants' predictions of future incidents by the company and their estimation of the necessity of a hostage, expecting to find a relationship to trustworthiness. If the public perceives a company to be trustworthy, they would then tend to trust them. As described earlier, trust was defined as a willingness to be vulnerable to the actions of another party, with the expectation that the party will not harm them, even though the party could do harm. Thus, the improvement in perceived trustworthiness of the company due to voluntary hostage posting would lead to the public's expectation that the company will prevent further incidents and will therefore not allow any further harm. We then set up the second hypothesis.

Hypothesis 2. Voluntary hostage posting will make the expectation of a future incident by the posters lower compared to imposed hostage posting.

At the same time, the improvement in perceived trustworthiness should lead the public to feel that the hostage is unnecessary because the company would be expected to act in the public interest, not solely in the company's interests. Thus, the public would judge the keeping of a hostage, which is a device to control the other party by virtue of the outcome structure, less necessary when the company posts a hostage voluntarily.

Hypothesis 3. Voluntary hostage posting will make the estimation of the necessity of keeping the hostage smaller compared to imposed hostage posting.

The influence of the initial evaluation

The main assertion of this article is that voluntary hostage posting restores perceived trustworthiness of a company or a government when it has been endangered by an incident. However, this raises another question: Is voluntary hostage posting effective for every truster? We would answer it is likely that the initial evaluation by trusters of the trustees would be affected by the manner of hostage posting. We come to this hypothesis from a series of empirical studies conducted by Yamagishi and Yamagishi (1994); Yamagishi, Kikuchi, and Kosugi (1999). First, they measured participants' initial level of trust of others, and then observed their response to information about trustworthiness of a target person. Contrary to the popular view (for example, Garske, 1976; Gurtman & Lion, 1982), results of both experiments, using a scenario procedure and the Prisoners Dilemma game, showed that participants whose initial level of trust is high are neither gullible nor credulous. Rather, results consistently showed that they pay more attention to the information potentially revealing untrustworthiness of a target person. Those whose initial level of trust is low are not sensitive to information concerning the trustworthiness of a target person. In the field of risk management, Cvetkovich et al. (2002) also found that participants' initial trust level affects the importance of new information in the evaluation of the trustworthiness of hazard managers.

According to the abovementioned studies, those who have a favorable initial attitude towards the company responsible for an incident are not simply insensitive to the company's wrongful behavior. Rather, they pay close attention to the company's behavior and to information regarding the trustworthiness of the target company. Thus, after an incident, their perceived trustworthiness of the company will change depending on whether or not the company posts a hostage voluntarily. If the company voluntarily posts a hostage, the perceived trustworthiness will be restored after the incident. If the company's posting of a hostage is not voluntary, the perceived trustworthiness will not be affected by the hostage posting and may in fact be worsened because of the incident. On the other hand, those whose initial attitude is negative are presumed to not be sensitive to the information on the company's trustworthiness. Thus, their perception of trustworthiness of a target company will remain negative after an incident whether or not the company posts a hostage voluntarily. In brief, the effects of the voluntariness of hostage posting on the perception of trustworthiness after an incident will be moderated by the prior attitude.

In the discussion above, we used the studies of Yamagishi et al. (1999) and Cvetkovich et al. (2002) to predict the interaction between initial evaluation and hostage posting on perceived trustworthiness. In the application, we assumed that the initial trust level noted in their studies may be compatible with the initial attitude. This compatibility has not been proved rigorously. However, attitude towards a company is a broad evaluation, and it presumably involves perceived trustworthiness. Thus, we assumed them to be compatible in the context of the present study. Of course, without evidence the assumed compatibility shows a limitation of the present studies, and it remains to be addressed in future research.

Hypothesis 4. The difference in perceived trustworthiness between voluntary posters and imposed posters will be larger for respondents whose initial attitude towards the posters is positive compared to those whose initial attitude to the posters is negative.

In the rest of this paper we empirically examine the signaling effects of voluntary hostage posting in a situation where trust in a company is endangered by an adverse event. Using two paper-and-pencil vignette experiments, in which actual company incidents were

adopted as material, respondents' estimations of the companies trustworthiness, the possibility of future incidents, and the necessity of keeping the hostage are compared between conditions of hostage posting to test these hypotheses.

Experiment 1

In Experiment 1, we measured participants' initial attitude toward a well known manufacturer of musical instruments. We then provided them with a vignette concerning an incident caused by the company. Participants read a real newspaper article that reported the incident, followed by the description of their hostage posting after the incident. Participants in the *voluntary* condition were told that the company offered a hostage (a monitoring and penalty rule for any future dishonest practices) of their own accord, whereas those in the *imposed* condition were told that the company agreed to post the same hostage after demands from stakeholders. Finally, participants responded to items comprising dependent variables including perceived trustworthiness, prediction of future incidents, and necessity of a hostage. In this experiment, participants changed their perception of trustworthiness of the company twice. The first change was caused by information about an incident caused by the company, and the second change was a result of information about their hostage posting. Given that the initial level of perceived trustworthiness of the company is similar between conditions by the random assignment of participants, the differences in trustworthiness scores after the hostage posting, if detected, can be attributed to the type of posting because the information about the incident is identical. A similar deterioration of trustworthiness as a result of the incident is assumed in both conditions. Also, the initial attitude level measured can be used to infer the similarity of the initial level of perceived trustworthiness between conditions. Checking participants' initial attitude towards the company is also important in order to see whether the company we used was adequate to test the effects of hostage posting. We tried to find a well known company with a neutral reputation among the participants. If the reputation was either extremely negative or positive it would be difficult to detect the effects of hostage posting because the extreme initial attitude would mask the effects of the manipulations. In this article we examine the effects of hostage posting with the premise that the initial attitude towards the target company was neutral, even though some degree of individual variance would be expected.

In this experiment we tried to examine the signaling effects of hostage posting, not the effects of an incentive structure changed by hostage posting leading to trustworthy behavior. To avoid the confluence of these two effects in the participants' evaluation of trustworthiness and their prediction of future incidents, the same termination of posting (the ending of the monitoring and sanction rule) is introduced at the end of the vignette in both conditions. By this termination, the impact of hostage posting on the incentive structure is temporary. Thus, the nature of posting a hostage becomes important through its signaling function and not through the resulting changes in the incentive structure, and the difference in scores of dependent variables can be attributed to the signaling effect caused by the voluntary posting.

Method

Participants and design

A total of 198 (93 women, 105 men) undergraduate students participated in the experiment. Their mean age was 20.8 ($SD = 4.50$). They were enrolled in psychology classes at four universities in Japan. Because there was no gender effect in either of the two experiments, we will not mention gender further. This experiment used a hostage posting condition (voluntary vs. imposed) between-subjects design. Participants were randomly assigned to one of the two cells.

Materials and procedure

Participants received a booklet which included a reprint of a newspaper article about the instrument manufacturer incident, an explanatory description of how they dealt with the incident, and questionnaire items with Likert scales. At the beginning of the first page, there were six items to measure attitudes towards the company prior to the manipulations. The second page contained an article and a description of what the company did. Questionnaire items and scales were listed on the third page. The 141-word newspaper article was authentic, about 15 cm × 3.5 cm in size, reporting that: (a) a famous instrument manufacturer announced the recall of their electric piano because of the risk of electric shock, (b) the target of the recall was a part contained in 1073 pianos which had already been shipped to retailers, and (c) they had examined their products and found a bad connection in the electric circuit which might result in electric shock.

A description of how the company dealt with the incident was put below the article on the second page. The descriptions differed according to the experimental conditions. In *the voluntary condition*, the company, in addition to examinations of and improvements to their manufacturing process, offered and carried out the following measures voluntarily: "The company announced they would accept a monitoring committee, which consisted of researchers in universities and consumer representatives, to inspect the factory on demand. The results of the inspection would be announced through mass media and company brochures. They also publicly

promised to close the factory if the committee detected any deception on the part of management." After this announcement, the committee inspected the factory repeatedly and found no problems in its operation. The results of the inspection were announced as the company promised. Consequently, the committee decided themselves to disband and terminated the inspections. In *the imposed condition*, the company was described as having accepted the same conditions after government and consumer groups' demands. Sound operation was confirmed repeatedly by the committee, and the consequent breakup of the committee and termination of the inspection were described just as in the voluntary condition.

In both the voluntary and imposed conditions the committee was dissolved and the inspection terminated, suggesting that the monitoring and the preparation of sanctions were over. These descriptions were added in order to return the hostage and to restore the incentive of the companies to lessen their efforts in management. If the hostage had remained posted and participants trusted the company, two effects of hostage posting would have been conflated, (a) to have participants recognize an increase in the company's incentive to *behave* honestly, (b) to have participants inferring the disposition of the company to *be* honest. Thus we set up the termination of the monitoring and preparation of sanctions to reduce the external incentives on the company to behave honestly.

Participants were instructed to first answer the six attitude items at the beginning, then to turn the page and read the article and the following description on the second page carefully. After reading them, participants rated 22 items about the company on Likert type items. After completing the questionnaire, the purpose of the study was explained to participants, they were thanked, and then dismissed.

Dependent measures

A five-point Likert scale was used for all items (see Appendix A), which ranged from "strongly disagree" to "strongly agree." Participants' ratings were quantified so that the larger value indicated positive evaluations. The mean of the items in each group was used as a composite to measure prior attitude and ending the necessity of a hostage. These composites had acceptable reliabilities (Cronbach's $\alpha = .78, .70$, respectively). For the estimation of future incident, Cronbach's α was the highest (.73) when item 17, "this company will not make a faulty product anymore" was excluded. Thus, the mean of the other five items was used as a composite of the future incident as perceived by participants. We assumed a two-factor model of trustworthiness, constructed from motivational factor and ability factor, and prepared five items for each factor, from item no. 7 to no. 11 for motivational factor and from item no. 12 to no. 16 for ability factor in Appendix A. These 10 items were factor analyzed using a principal components model with an oblique rotation. Two factors were extracted as having eigenvalue greater than 1.0. Factor 1 with an eigenvalue of 4.48 accounted for 44.8% of the variance, and Factor 2 with an eigenvalue of 1.37 accounted for 13.6% of the variance. No other factors had eigenvalue exceeding 1.0. The correlation between Factor 1 and Factor 2 was .506. All five items for motivational factor showed high item loadings for rotated Factor 1 (.50 or greater), with two items prepared for ability factor (item no. 12 of .71 and item no. 13 of .84). Item loadings for rotated Factor 2 were higher (.70 or greater) only in the other three items prepared for ability factor. We then used the mean of the former seven items (from item no. 7 to no. 13) as a motivational composite and the mean of the other three items (from item no. 14 to no. 16) as an ability composite. Their Cronbach's αs were .84 and .78, respectively.

Results and discussion

Data from five participants were excluded due to missing responses. Data were therefore analyzed for 193 participants. The mean scores of prior attitude in the voluntary condition and the imposed condition were 3.11 ($SD = .51$) and 3.05 ($SD = .44$), respectively. These scores were close to each other, and a t test (voluntary vs. imposed) on the pretest scores found no significant difference, as expected ($t(191) = .79$, $p > .43$, $d = .13$). This result suggests random assignment of participants to the conditions made prior attitude toward the company equal between the experimental conditions.

The mean scores for other composites are given in Table 1. In the posttest, motivation and ability scores in the voluntary condition, as expected, were higher than those in the imposed condition. The t test on the two scores yielded the expected significant difference between conditions ($t(191) = 12.58$, $d = 1.33$, $t(191) = 5.70$, $d = .76$, respectively, $ps < .001$). These results suggest that the perceived trustworthiness of the companies rose more as a result of voluntary hostage posting than imposed hostage posting. These findings provided support for Hypothesis 1.

Table 1
Mean scores of each dependent measure in voluntary and imposed conditions in Experiment 1

Measure	Condition	
	Voluntary ($n = 96$)	Imposed ($n = 97$)
Motivation	3.81 (.65)	2.77 (.50)
Ability	3.54 (.73)	2.94 (.72)
Future incidents	3.43 (.74)	2.85 (.46)
Ending necessity of hostage	3.45 (.70)	3.16 (.61)

Note. Values in parentheses are standard deviations. All measures are based on Likert scales ranging from 1 to 5, with larger values indicating positive evaluations.

Hypothesis 2 concerns the estimation of future incidents. The results of the t test on future incidents were in the same direction as the perceived trustworthiness described above, $t(191)=6.50$, $p<.001$, $d=.85$. The score in the voluntary condition exceeded that of the imposed condition. This result provided support for Hypothesis 2.

Hypothesis 3 predicted that voluntary hostage posting would induce participants to estimate the hostage as being less necessary than it would be with imposed hostage posting. This is because voluntary hostage posting would raise the components of trustworthiness (perceived motivation and ability) more than imposed hostage posting, which would not raise the components, as demonstrated in Hypothesis 1. The score for the hostage being unnecessary in the voluntary condition was higher than that in the imposed condition, providing support for Hypothesis 3. A t test showed a significant difference in scores between the voluntary and the imposed conditions ($t(190)=3.15$, $p<.01$, $d=.43$). This finding suggests that the public would be less likely to demand a further hostage from the organization that posted it voluntarily than they would of the organization that was perceived to be reluctant to provide it.

Hypothesis 4 concerns the interaction between prior attitude and hostage posting on perceived trustworthiness. To test this hypothesis, we sorted participants depending on their scores of prior attitude towards the instrument manufacturer and divided them into four groups, selecting the highest scores as a positive attitude group and the lowest scores as a negative attitude group. The mean scores of the positive and the negative attitude group were 3.74 ($SD=.24$, $n=43$) and 2.48 ($SD=.24$, $n=47$), respectively ($t(88)=24.94$, $p<.001$, $d=1.85$). Fig. 1 shows the mean scores of the motivation composite and the ability composite in each posting condition by attitude group. In both composites, differences in scores between voluntary and imposed posting conditions were larger in positive groups than those in the negative groups, as expected. The results of the analysis of variance on the scores showed significant prior attitude × posting condition interactions (motivation, $F(1, 86)=6.89$, $p<.01$, partial $\eta^2=.074$; ability, $F(1, 86)=12.21$, $p<.001$, partial $\eta^2=.124$). These findings indicate that prior attitude towards the company influenced the effects of hostage posting, suggesting that the respondents whose attitude towards the company were positive were more sensitive to how the hostage was posted by the company. These results provided support for Hypothesis 4. ANOVAs also yielded significant main effects for posting conditions in all composites (motivation, $F(1, 86)=61.89$, partial $\eta^2=.418$; ability, $F(1, 86)=14.99$, partial $\eta^2=.148$, $ps<.001$), consistent to the results of t tests. No main effects for prior attitude were significant (motivation, $F(1, 86)=.84$, $p>.36$, partial $\eta^2=.010$; ability, $F(1, 86)=.14$, $p>.70$, partial $\eta^2=.002$).

To sum up, the results of Experiment 1 provided initial support for our hypothesis that voluntary hostage posting after an incident would raise the perceived trustworthiness of an organization responsible for an incident. Results also confirmed that the expectation of future incidents by the organization and the estimation of the necessity of a hostage would decrease due to voluntary posting. The findings of interaction between prior attitude and hostage posting indicated that voluntary posting after an incident is important, especially for the public whose prior attitudes toward the organization are positive.

On the other hand, the design of this experiment suggests a limitation. There were three points in time when participants' changing evaluation of trustworthiness could be measured, the initial point, the point after reading about the incident, and the point after reading about the hostage posting. In Experiment 1, only the final point was utilized to assess the effects of hostage posting. Because of this, we cannot confirm how the perceived trustworthiness would be restored compared to that of the initial level, because that was not directly measured. Though the evaluations of voluntary posters by the respondents were better than those for imposed posters, it was unclear whether the evaluations for voluntary posters increased or those of imposed posters decreased as a result of the respondents' perceived trustworthiness

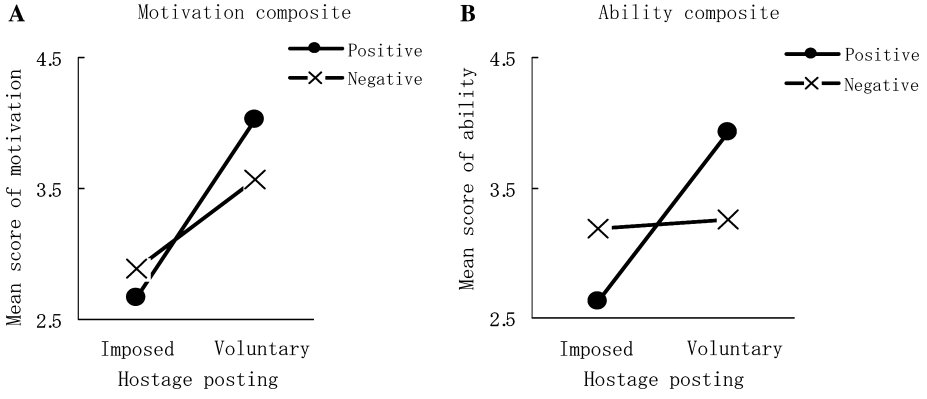

Fig. 1. Interaction between prior attitude and hostage posting on perceived trustworthiness in Experiment 1.

level after the posting. The procedure in Experiment 1 could not address the questions of whether either voluntary or imposed hostage posting might have lessened the perceived trustworthiness more than had there been no hostage posting. We conducted Experiment 2 to amend the design and procedure so that these limitations in Experiment 1 could be addressed.

Experiment 2

Experiment 2 was conducted to confirm the replication of the results of Experiment 1, improving on its design. In Experiment 2 participants were provided twice with identical items regarding perceived trustworthiness, before and after they read the vignette. The addition of the prior measurement to the incident information made it possible to assess the initial level of perceived trustworthiness and the degree of restoration due to hostage posting compared to the initial level. In addition, we added a third condition (a control condition) to measure perceived trustworthiness after respondents had read about the incident with no mention of hostage posting. The degree of deterioration of perceived trustworthiness caused by the incident could be assessed by the within comparison of the scores between the initial trustworthiness and those after the incident in the control condition. Furthermore, comparisons of trustworthiness scores between the voluntary and control conditions, and between the imposed and control conditions, made it possible to assess the degree of restoration of perceived trustworthiness due to the hostage from the deteriorated level prior to the hostage being offered. Another difference between Experiment 2 and Experiment 1 was that the material in the vignette was an issue that was ongoing and familiar to the participants. In the first half of 2004, the expansion of the bird flu infection was one of the largest public health concerns in Asian countries. In Experiment 2, a famous fast food company found to be using chicken products from a country affected by bird flu was used as a trustee in the vignette. As described earlier, the dependent variables regarding the motivational factor of trustworthiness were also expanded in order to generalize the findings from Experiment 1.

Method

Participants and design

A total of 313 (134 women and 179 men) undergraduate students participated in Experiment 2. Their mean age was 20.1 ($SD = 2.10$). They were enrolled in psychology classes at two universities in Japan. This experiment used a hostage posting condition (voluntary vs. imposed vs. control) between-subjects. Participants were randomly assigned to one of the three cells.

Materials and procedure

Two reprints of real newspaper articles, one of 182 words, about 10 cm × 15 cm in size and one of 92 words, about 6 cm × 7 cm in size, were used in Experiment 2. The first article reported that: (a) an outbreak of bird flu was suspected in Thailand and (b) the Japanese government had banned the importation of chicken from Thailand. The second article reported that: (c) the government of Thailand confirmed the bird flu infection in their citizens and (d) had confirmed five deaths of six infected, and nine other deaths in 21 suspected cases. Two sets of descriptions were placed below these articles. The upper set showed the name of a famous fast food restaurant and said that they had been found using chicken from Thailand. The lower set described how the company dealt with the incident. The contents differed depending on the conditions (voluntary vs. imposed vs. control). In *the voluntary condition*, the company, in addition to changing the origin of the chicken they used, voluntarily offered and carried out the same measures as described in Experiment 1. When inspections by the committee found no problems, the committee decided to disband and terminated the inspections, as in Experiment 1. In *the imposed condition*, the company was described as having accepted the same conditions after government and consumer groups' demands. The return to sound operation, the consequent break-up of the committee and termination of the inspection were described just as in the voluntary condition. In *the control condition*, there was no information about the company's hostage posting, only the change in the origin of the chicken, as in other two conditions.

At the beginning of testing, participants were instructed to answer the questions related to their attitude towards, trustworthiness of, and experience with the target restaurant. After the first questionnaire was retrieved, participants were provided with a vignette with the second questionnaire. After reading about the incident and measures taken by the company, participants rated 30 items regarding their estimation of trustworthiness, the possibility of future incidents, the necessity of a hostage, and their intention regarding future usage of the restaurant. After completing the questionnaire, participants were briefed, thanked, and dismissed.

Dependent measures

Participants' experience of the target restaurant, frequency of usage of, and knowledge about the restaurant were measured by five-point scales. The scales of frequency were "not at all," "not frequently," "sometimes," "frequently," and "very frequently." The scales of knowledge were "not at all," "very little," "somewhat," "fairly," and "extremely." A five-point Likert scale was used for all the other items, which ranged from "strongly disagree" to "strongly agree," as in Experiment 1.

We added items to measure participants' intention of future usage of the restaurant after they had read the vignette. These items are shown in Appendix B. Items to measure prior attitude, estimation of the likelihood of future incidents, and ending the necessity of a hostage were slightly modified to account for the change of issue from risk in a piano in Experiment 1 to risk in food in this experiment. Prior attitude was measured by five items. They were 1, 2, 5, and 6 in Appendix A, and "I would like to eat their product." The estimation of future incidents was measured by five items, eliminating item 21 from the future incident composite in Appendix A. The estimation of ending the necessity of a hostage was measured by items 23 and 24 (with the change of allowance of "free access to their information" to "surprise inspections"), 25 and 27 in Appendix A, and "the authorities concerned should not ease their supervision of this company." Cronbach's αs of prior attitude, estimation of future incidents, estimation of ending the necessity of a hostage, and intention of usage of the restaurant were .86, .84, .83, and .90, respectively. As these indexes of reliability were acceptable, the means of the items were used as a composite as in the previous experiment.

In Experiment 2, we assumed a three-factor model of trustworthiness constructed from the integrity, benevolence, and ability factors. Thus, we added items to measure the motivational factor of trustworthiness, dividing it into an integrity aspect and a benevolence aspect (Appendix B). The five items for an ability factor were the same as in Experiment 1 (item no. 12–no. 16 in Appendix A). These 15 items were factor analyzed using a principal components model with an oblique rotation. Again, two factors were extracted as having eigenvalue greater than 1.0. Factor 1 with an eigenvalue of 8.05 accounted for 53.7% of the variance, and Factor 2 with an eigenvalue of 1.51 accounted for 10.0% of the variance. No other factors had eigenvalue exceeding 1.0. Correlation between Factor 1 and Factor 2 was .568. All the 10 items prepared for integrity and benevolence showed high item loadings for rotated Factor 1 (.50 or greater) while the other five items for ability showed high loadings for rotated Factor 2 (.35 or greater). We then used the mean of the former 10 items as a motivational composite and the mean of the other five items as an ability composite. Their Cronbach's αs were .94 and .80, respectively.

Results and discussion

Data from six participants were excluded due to missing responses. Data were therefore analyzed for 307 participants. The ratio of responses to their frequency of the target restaurant usage was .075 for "not at all," .401 for "not frequently," .481 for "sometimes," .033 for "frequently," and .010 for "very frequently." The ratio of responses to the knowledge scale was .003 for "not at all," .055 for "very little," .730 for "somewhat," .144 for "fairly," and .068 for "extremely." These results suggested that the target restaurant was familiar to participants in general. The mean scores of prior attitude in the voluntary condition, imposed condition, and control condition were 3.38 ($SD = .90$), 3.43 ($SD = .73$), and 3.40 ($SD = .88$), respectively. An ANOVA (voluntary vs. imposed vs. control) on the prior attitude scores found no significant effect, as expected ($F(2, 303) = .10$, $p > .90$, $\eta^2 = .001$). These results suggest that random assignment to the conditions was successful in making three equal groups of participants concerning their prior attitude towards the target restaurant.

The mean scores of the composite variables concerning trustworthiness after reading the vignette are given in Table 2. Both the motivation and the ability scores in the voluntary condition exceeded those in the imposed and the control conditions. In addition, scores in the voluntary condition were greater than 3, which was the neutral point of the scales. On the other hand, scores in the imposed and the control conditions were less than 3. As we formulated hypotheses prior to the analysis, planned comparisons using multiple t tests were performed to test the hypothesis. The results of multiple t tests on scores of the motivation composite yielded significant differences between the voluntary and the imposed condition ($t(303) = 6.01$, $p < .001$, $d = .80$), between the voluntary and the control conditions ($t(303) = 7.75$, $p < .001$, $d = .99$), but not between the imposed and the control condition ($t(303) = 1.61$, *n.s.*), as expected. Comparisons of mean scores of perceived ability yielded a marginally significant difference between the voluntary and the imposed condition ($t(300) = 1.88$, $p < .10$, $d = .27$), a significant difference between the voluntary and the control condition ($t(300) = 2.88$, $p < .01$, $d = .41$), but not between the imposed and the control condition ($t(300) = 1.11$, *n.s.*, $d = .14$). In sum, these findings replicated the results of Experiment 1 for the most part, providing support for Hypothesis 1.

The pretest–posttest differences in scores for each composite were calculated by taking pretest scores from

Table 2
Mean scores of each dependent measure in voluntary, imposed, and control conditions in Experiment 2

Measure	Condition		
	Voluntary ($n = 106$)	Imposed ($n = 97$)	Control ($n = 104$)
Motivation	3.35 (.85)	2.65 (.78)	2.48 (.77)
Ability	3.18 (.75)	2.99 (.65)	2.89 (.71)
Future incidents	3.12 (.79)	3.02 (.71)	3.01 (.75)
Ending necessity of hostage	2.83 (.91)	2.72 (.77)	2.45 (.78)
Intention of usage	3.73 (.97)	3.71 (.82)	3.33 (.94)

Note. Values in parentheses are standard deviations. All measures are based on Likert scales ranging from 1 to 5, with larger values indicating positive evaluations.

Table 3
Pretest–posttest differences in scores for trustworthiness measures in voluntary, imposed, and control conditions in Experiment 2

Measure	Condition		
	Voluntary ($n = 106$)	Imposed ($n = 97$)	Control ($n = 104$)
Motivation	−.02 (.85)	−.70 (.85)	−.86 (.88)
Ability	−.12 (.70)	−.37 (.72)	−.45 (.73)

Note. Values in parentheses are standard deviations. Larger negative values indicate larger deterioration in evaluations.

posttest scores. As seen in Table 3, values in the voluntary condition were close to zero. This finding indicates that voluntary hostage posting recovered perceived trustworthiness of the company nearly to the initial level. On the other hand, values in the imposed condition were close to those in the control condition that posted no hostage. This suggests that posting a hostage after demands by stakeholders had no effect on recovering perceived trustworthiness. One-way ANOVAs on difference values for each composite were performed. The main effects for posting conditions were significant in both composites (motivation, $F(2, 304) = 30.92$, $p < .001$, $\eta^2 = .169$; ability, $F(2, 300) = 5.64$, $p < .01$, $\eta^2 = .036$). Scheffe's multiple comparison tests for the difference values of composites showed that the values in the voluntary condition were significantly smaller than those in the other two conditions (motivation, $ps < .001$; ability, voluntary-imposed at $p < .05$, voluntary-control at $p < .01$) and that there were no significant differences in the values between the imposed condition and the control condition. Again, note that the incident was identical through the conditions, and the hostage of the imposed condition was identical to that of the voluntary condition. Therefore, the results of analysis indicate that the perceived trustworthiness of voluntary posters deteriorated to a level similar to the control condition, then recovered nearly to the initial level as a result of voluntary hostage posting. On the other hand, perceived trustworthiness of imposed posters remained deteriorated after the hostage was posted involuntary.

As seen in Table 2, participants' estimations regarding the possibility of future incidents did not differ among conditions. A multiple t test on scores for estimation of future incidents found no significant effect (voluntary–imposed, $t(300) = .90$, $d = .13$; voluntary–control, $t(300) = 1.01$, $d = .15$; imposed–control, $t(300) = .12$, $d = .01$, all *n.s.*). These results did not provide support for Hypothesis 2. Similarly, a multiple t test showed that scores of ending the necessity of a hostage between the voluntary and the imposed conditions were not significantly different ($t(301) = .90$, *n.s.*, $d = .13$). Differences in the scores between the voluntary and the control condition, and those between the imposed and the control condition were significant ($t(301) = 3.32$, $p < .01$, $d = .45$, $t(301) = 2.49$, $p < .05$, $d = .32$, respectively). Though participants estimated that the hostage was less necessary in the voluntary condition compared to the control condition in which no hostage was posted, they did not estimate the hostage to be less necessary when it was posted voluntarily than when the posting was imposed. This finding did not provide support for Hypothesis 3.

As in Experiment 1, the participants in the voluntary and the imposed conditions were sorted and divided into four groups depending on their scores of the prior attitude composite, then those with the highest and the lowest in the score were extracted as the positive attitude group and the negative attitude group, in order to examine the interaction of the posting style with prior attitude on the perceived trustworthiness. Data from the control condition were eliminated in this analysis because participants read no information on hostage posting. The mean scores of the positive and the negative attitude group were 4.36 ($SD = .34$, $n = 49$) and 2.28 ($SD = .50$, $n = 51$), respectively ($t(98) = 24.33$, $p < .001$, $d = 1.84$). Fig. 2 shows the mean scores of the motivation and ability composite in each posting condition by attitude group. The differences in scores of the composite between the voluntary and imposed posting condition were larger in positive groups than those in the negative groups. These results for the most part replicated those

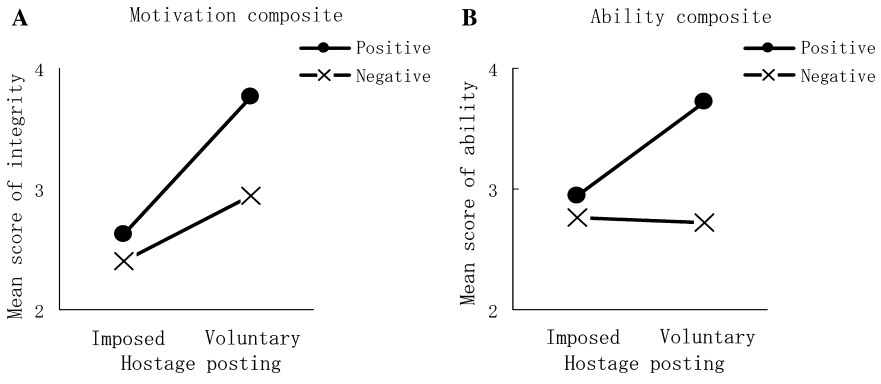

Fig. 2. Interaction between prior attitude and hostage posting on perceived trustworthiness in Experiment 2.

in Experiment 1. The results of ANOVA on the four scores showed marginally significant prior attitude × posting condition interaction for the motivation composite, and significant interaction for ability composite (motivation, $F(1, 96) = 3.38$, $p < .07$, partial $\eta^2 = .034$; ability, $F(1, 95) = 6.20$, $p < .05$, partial $\eta^2 = .061$). These findings suggest that the respondents whose prior attitude about the restaurant were positive would likely be more sensitive to how the hostage was posted in their estimations of trustworthiness of the restaurant, providing support for Hypothesis 4. ANOVAs also yielded significant main effects for the posting condition (motivation, $F(1, 96) = 26.81$, $p < .001$, partial $\eta^2 = .218$; ability, $F(1, 95) = 4.44$, $p < .05$, partial $\eta^2 = .044$). In both composites, main effects for prior attitude were significant (motivation, $F(1, 96) = 10.32$, $p < .01$, partial $\eta^2 = .097$; ability, $F(1, 95) = 14.31$, $p < .001$, partial $\eta^2 = .130$).

In brief, Experiment 2 largely replicated the results of Experiment 1 for the perception of trustworthiness of the trustees, providing supporting evidence for Hypothesis 1 and Hypothesis 4. For the estimation of the possibility of future incidents and ending the necessity of keeping a hostage, however, the results of this experiment did not replicate those in Experiment 1, which provided support for Hypothesis 2 and Hypothesis 3. We will discuss this inconsistency in the General discussion.

The values in the bottom row in Table 2 show the mean scores of the respondents' intention of future usage of the restaurant by condition. It was expected that the score in the voluntary condition would exceed that in the imposed condition, because respondents would perceive a higher trustworthiness in the former condition. The results shown in the table, however, suggest that the future usage intentions were at the same level in the voluntary and in the imposed conditions. A multiple t test showed that scores of future usage intention between the voluntary and the imposed conditions were not significantly different ($t(298) = .1.71$, *n.s.*, $d = .02$). Differences in the scores between the voluntary and the control condition, and those between the imposed and the control condition were significant ($t(298) = 3.01$, $p < .01$, $d = .43$, $t(298) = 3.02$, $p < .01$, $d = .41$, respectively). Why did the usage intention in the voluntary condition not exceed that in the imposed condition? There seem to be two interpretations. One is that there were several factors besides trust in the trustees, for example, participants' taste in food, transportation to the shop, and so on, and that these affected the behavioral intention sufficiently that the effect of voluntary posting was difficult to detect. The other one is that the signaling effect of voluntary hostage posting substantially stays within the *perception* of trustworthiness and does not affect people's *intention* to rely on the hostage posters. If the latter interpretation is true, the meaning of the model of voluntary hostage posting would be extremely limited in its application for management practice as well as for the theoretical development of trust. Thus, we decided to conduct an additional experiment as Experiment 3 in order to confirm whether the voluntary posting affected trust at the behavioral level, going beyond the paper-and-pencil estimation of a vignette.

Experiment 3

We conducted an interpersonal trust game to examine the effects of voluntary hostage posting at the behavioral level. Interpersonal game experiments have other two benefits: (a) to confirm posting effects where the participants' interests are involved, which is difficult to examine using a paper-and-pencil method and (b) to eliminate the effects of pretest factors like participants' tastes and prior attitude towards the trustees, which are inevitably involved in a vignette experiment using real companies.

Method

Game

We used the two-person trust game shown in Fig. 3. One of the two players was actually an experimenter, but the real participant did not know that. At the beginning of the experiment, each participant was given 500 yen (about $5) as a reward for participation. They were then asked to play a one-shot game with the given money. In this game, Player 1 first decided how much of her/his own money s/he would give to Player 2. The amount that could be given was from 0 to 500 in 100 yen increments. Player 2 had the opportunity to increase the given money. Player 2 left the room, went to the other lab, and asked the staff to increase the money. In this game, the amount could be tripled. If Player 1 gave all 500 yen, Player 2 would have 1500 yen in addition to 500 yen. After Player 2 had received the increased amount of money, she had a choice about whether or not to return to the room where Player 1 was waiting. If she came back, the amount of money would be split equally. In this case, each would earn 1000 yen. If she did not, she could go leave with 2000 yen. When she went to the other lab, she was asked to take all her belongings so that there was no incentive to return. In this situation, the critical factor in Player 1's decision was how much Player 1 trusted Player 2. Thus, the amount of submitted money by Player 1 was the measure of trust of Player 2.

Participants

Forty-four Japanese undergraduate students (27 women and 17 men) were independently recruited. Monetary rewards for participation were emphasized in recruiting.

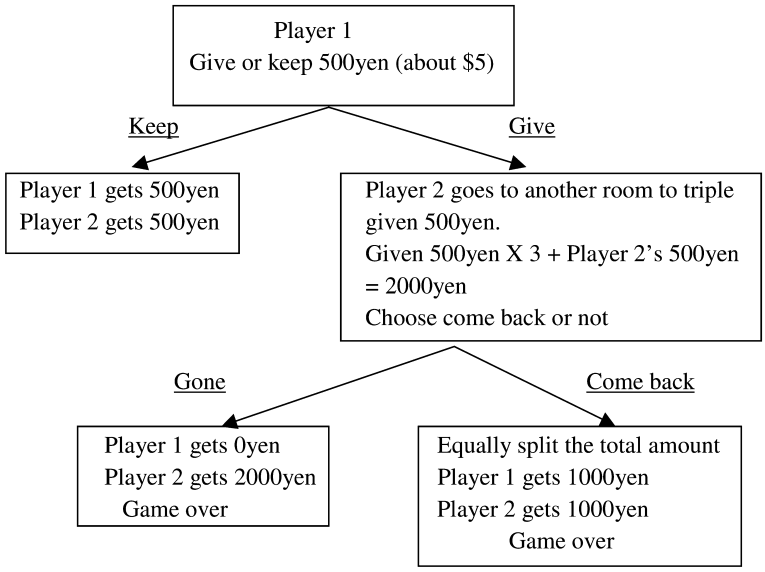

Fig. 3. The scheme of the trust game in Experiment 3. This example shows an extreme case. Player 1 actually chose the amount between 0 and 500yen by 100yen.

Design

A two factorial design was used. Hostage posting types were between subjects (imposed vs. voluntary) and games (1st and 2nd) were within subjects.

Procedure

Two participants, one of them an experimenter, separately entered the experiment room, which contained two desks and chairs. There was a partition between the desks so that participants could not see or communicate with each other. Another experimenter told them to read the instructions shown on the computer display. The instructions are shown in Fig. 3. The experimenter then asked a few questions to confirm that the real participant clearly understood the game. No participant failed in the experiment. Next, the experimenter asked them to draw a card from a set of two cards to assign their roles. The cards were artificially arranged so that the role of Player 1 was assigned to the real participant. After the drawing, the experimenter made sure that they understood every procedure and did an experimental manipulation. Either of the two types of hostage were provided as described below. In the imposed condition, the experimenter told the fake participant (the other experimenter) to leave his wallet and cellular phone in the experiment room. On the other hand, in the voluntary condition, the fake participant voluntarily offered to leave these items in the room. Before running the experiment, we had confirmed that a wallet and a cellular phone were more valuable to undergraduate students than the maximum profit of this experiment (2000 yen, about $20). For this manipulation, it was not a rational choice to leave after he had tripled the money. The monologue by the experimenter and fake participant was exactly the same in both conditions. The same hostage was provided in both conditions. The only difference between the two conditions was whether or not the fake participant offered to leave the valuable items voluntarily. After the experiment, the experimenter confirmed that the real participant understood the hostage provided.

Then, the experimenter asked Player 1 (the real participant) to decide how much of the 500 yen he wanted to give after the above manipulation. Player 1 put the money in a thick envelope so that the experimenter did not know how much he put in. Next, Player 2 (the fake participant) left the room with the envelope and pretended to go to another experiment room. During that time, the experimenter asked Player 1 to answer several questions including how trustworthy Player 2 was. After Player 1 had answered, the experimenter made a phone call to his supervisor to confirm the procedure. At that time, she acted as if she had made a mistake in the experiment. She came to Player 1 and explained that it was her mistake to ask Player 2 to leave his wallet and cellular phone, and asked to start the experiment again without leaving the items. The experimenter also said that the decision by Player 1 did not count. This procedure corresponds to the termination of inspection and sanction preparation in Experiment 1 and Experiment 2, and was necessary to examine how posting a hostage in the current game affected Player 1's decision in the next game after removing hostage.

After Player 2 returned, the experimenter stated that they were starting the game over. The experimenter emphasized that this time Player 2 brought the given money and all her belongings including her wallet and cellular phone. Player 1 then decided on the amount of the money and answered the same set of questions as in the previous game. Since Player 2 was a fake participant,

she came back to the room and they split their profit. Finally, Player 1 answered a postquestionnaire and received the money. It took 45 min to complete the experiment. After all 44 participants completed the experiment, a debriefing session was held.

Results and discussion

Since one participant did not remember the manipulation of the hostage posting, 43 participants were used in the following analyses.

Perceived trustworthiness in Player 2 after game 1

We asked the participant how trustworthy their partner was in a Likert type scale from 1 (not at all) to 9 (completely). Consistent with the results in Experiment 1 and Experiment 2, the mean score of trustworthiness was significantly higher in the voluntary condition than in the imposed condition (voluntary, 5.32; imposed, 4.20; $t(40) = 2.44$, $p < .05$, $d = .71$). This result proved consistent with the previous experiments using the questionnaire and also showed the validity of the experimental manipulation.

Amount of submitted money

Fig. 4 shows the amount of money in each game. In game 1, the amount of money in each posting condition was about the same. In game 2, the amount increased in the voluntary condition regardless of there being no hostage. On the other hand, the amount decreased in the imposed condition. The results of ANOVA on the amount of money showed that the interaction effect of posting conditions and games was significant ($F(1, 41) = 4.21$, $p < .05$, partial $\eta^2 = .093$).

As expected, participants were more likely to trust and depend on others when the hostage was provided voluntarily, whereas distrust was higher when the hostage was provoked by an external authority. In sum, the findings in Experiment 3 suggest a voluntary hostage posting effect not only on the trusters' perception of trustworthiness of trustees but also on their actual behavior, where the trusters risked their interests.

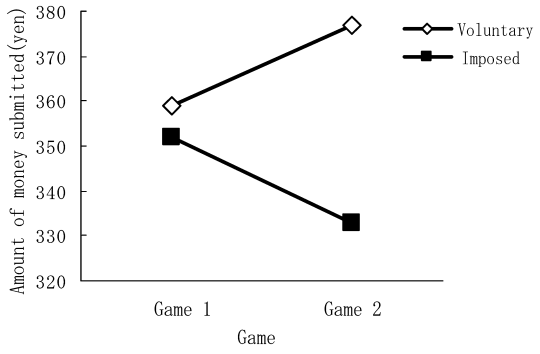

Fig. 4. The amount of submitted money in each game by condition in Experiment 3.

The voluntary posters in this experiment offered to leave their valuable items and did not offer any monitoring of themselves as in Experiment 1 and Experiment 2. However, the provision of the monitoring was not needed in this experiment because the valuables would be forfeited if the trustees did not return. The penalty could be given without continuous monitoring of the trustees' deception. On the other hand, monitoring was needed in the previous two experiments in order to detect the trustees' deception and to give them the penalty. Therefore, all manipulations in these three experiments can be regarded as a temporary provision of a hostage.

General discussion

The central thesis of this research was that voluntary hostage posting—the provision of monitoring and sanctions in the case of future dishonest practices—by trustees would raise trusters' perception of trustworthiness in them. Comprehensive trustworthiness perceived in a company or a government is composed of the public's expectation of their motivation to behave in a trustworthy manner and their ability to resolve the problem concerned. We hypothesized that voluntary hostage posting by organizations would function as a signal of their trustworthiness and enhance public estimation of their motivation and ability, and thus the comprehensive trustworthiness of them. On the other hand, imposed hostage posting, even though it induces the same incentive structure, would not enhance the perception of trustworthiness because the public attributes the cooperative behavior to factors external to the character of the management. While imposed hostage posters would be regarded as not deceiving the public only in order to avoid sanctions, voluntary hostage posters would be regarded as not deceiving the public because of their trustworthy disposition. Thus, the public would trust the voluntary posters even after the hostage system is terminated. The results of the three experiments provided support for our hypotheses as a whole, demonstrating the signaling effects of voluntary hostage posting on the enhancement of perceived trustworthiness.

We used dependent items to measure perceived trustworthiness assuming a two-factor model of trustworthiness—ability and motivation—in Experiment 1, and a three-factor model—ability, integrity, and benevolence—in Experiment 2. The results of the factor analysis in Experiment 1 and Experiment 2 consistently supported the two-factor model, failing to distinguish integrity from benevolence. Why was benevolence in this study not distinguished from integrity in spite of the fact that the three-factor model of trustworthiness has been confirmed in previous studies (Mayer & Davis, 1999; Mayer et al., 1995)? One reason might be the limited

information available to respondents in the vignette. If participants had estimated the trustworthiness of someone whom they observe daily, like co-workers, bosses or friends, the two sub-components may have been more distinguishable because more information would have been available. In this study, however, the information available in the vignette was so limited that it might have been difficult for respondents to estimate trustworthiness in a detailed way. Nevertheless, adopting the two-factor model, the reliability of motivation and ability composite was high enough, and the items in the motivation composite reflected the concepts of both integrity and benevolence, as seen in Appendix B. Therefore, we believe that the results of our experiments confirmed the thesis that voluntary hostage posting recovers comprehensive trustworthiness of trustees, but that the passive acceptance of it does not. It remains for further research to determine whether voluntary posting affects sub-components of trustworthiness equally, or whether it affects sub-components differently.

The monitoring of organizations has costs and requires resources. For example, monitoring the operation of factories requires outside observers' time and money. In Experiment 1, participants in voluntary conditions evaluated a future incident as less likely to recur and believed that therefore there would be less need for a voluntary hostage. This suggests that voluntary hostage posting may be able to reduce future transaction costs for keeping a hostage. Fukuyama (1995) and Putnam (1993) explained how trust improves social efficiency by reducing costs for economic transactions. The findings in Experiment 1 suggest that posting a hostage voluntarily is one of the ways to reduce the cost of social transaction because it enhances trust. The results concerning estimations of future incidents and ending the necessity of a hostage in Experiment 2, however, did not show the same effects due to voluntary posting. Why were there differences between Experiment 1 and Experiment 2? One reason may be the difference in the locus of origin of the incidents in the two vignettes. The origin of the incident used in Experiment 1 (instrument manufacturer) was in the factory of the target company. On the other hand, bird flu, which was the fundamental origin of the incident used in Experiment 2 (fast food restaurant), was not caused by the target company. If the origin had been in the company, the recovery of competence and positive motivation for consumers—thus, trustworthiness—would have directly reduced the perceived possibility of a recurrence of an incident in the future. On the other hand, if the fundamental origin is outside of the company, the risk of a further incident cannot be sufficiently controlled however much good faith effort the company might make. These differences may have caused the estimated possibility of future incidents in the voluntary condition to not significantly exceed that in the imposed condition, despite the fact that the perceived trustworthiness of the company in the former condition was much more positive than in the latter condition. And for that reason, participants in the voluntary condition might judge monitoring to be necessary at a level similar to that in the imposed condition. This interpretation of the results of Experiment 2 is no more than a post hoc explanation, and future empirical studies putting the locus of origin of the adverse event into independent variables are needed to confirm this interpretation. Results of Experiment 2, however, might be meaningful, suggesting that voluntary hostage posting is effective in improving the perceived trustworthiness even when the adverse event is not the fault of the trustee.

This research examined the effects of voluntary hostage posting mainly in the context of endangered trust of companies after incidents. However, the model of hostage posting as a device to resolve conflict is a general one (Gautschi, 1999; Keren & Raub, 1993; Mlicki, 1996; Raub & Keren, 1993) and would not be limited only to companies' or governments' risk management situations. Whatever the combination of actors, for example, consumer and manufacturer, citizen and government, or interpersonal relations, hostage posting reduces the incentive to deceive the other. This means that hostage posting reduces social uncertainty, and increases the other's expectation that the poster will not deceive. This research extends the hostage model theoretically and empirically, claiming that posting a hostage will have greater effects on resolving conflict when it is done voluntarily. The results of Experiment 3, which we conducted as an additional experiment to examine whether the voluntary posting effect would go beyond paper-and-pencil stimuli, confirmed its influences on the actual behavior in an interpersonal exchange where real interests existed. To further confirm the generalizability of the effects of voluntary hostage posting, future research should be extended to other areas where social uncertainty exists. Participants other than university students should also be recruited in future research to confirm generalizability.

Implications for practice to restore public trust

As we described at the beginning of this paper, public trust in the policy makers is one of the major determinants of whether their policy is successful or not. We interpreted the results of a survey by Slovic (1993), which suggested that the best way to increase public trust was to delegate to the public the authority to monitor and shut down a nuclear plant, as showing the effect of hostage posting in a conflict situation. We also hypothesized that the essence of the effect is in the voluntariness of the behavior and tested it in the context of endangered trust of companies after adverse events. The results of our studies suggest that agents responsible should provide a monitoring and sanction system for themselves *voluntarily*. In particular, agents who would like to recover public trust after an inci-

dent should post a hostage before stakeholders demand it. Government authorities sometimes impose new regulations on industries with the intention not only of protecting citizens' safety but also in order to recover trust. For example, governments regulate the management of repository sites for nuclear waste, GMO foods, and the import of chicken from countries affected by bird flu, in order to recover public trust, and industries generally abide by the regulations. Our research suggests that government regulation and acceptance of that regulation by industries will not improve the public perception of trustworthiness in the industries. So, a question arises concerning the effects of imposition of regulations. Does it deteriorate the public perception of trustworthiness in industries more than in a situation where no regulations are imposed? The results of Experiment 2 suggested no significant differences on scores of trustworthiness between the imposed and the control condition, indicating that imposition of regulation did not harm respondents' perceived trustworthiness. In sum, this research, at least, found that the imposition of regulations had no effects on *improvement* of trustworthiness, and that accepting imposed regulations was less effective than self-regulation. However, we must reserve a final conclusion that the imposition of regulation has no negative impact on trust because the evidence was drawn from only a single experiment (Experiment 2). Here again, future research is needed to confirm whether or not the imposition of hostage posting deteriorates trust.

An advantage of hostage posting as a device to recover trust is that it could create an opportunity for the posters themselves. Regardless of what companies or governmental agents do to improve the quality of their products or services after incidents, they will not recover trust unless the public selects their products or services and finds the quality to be sufficiently improved. However, the public might be unwilling to select a company or organization that has previously caused an incident. For example, consumers likely would not purchase a brand of food that has had a food poisoning scandal. Thus, improvement of safety, without any active approaches to the public, may not increase trust. With voluntary hostage posting, however, they can control when to begin recovery, and a voluntary declaration to post a hostage can be made at the companies' or agents' will.

Finally, we would like to caution every organization concerned about public trust not to use voluntary hostage posting as a cheap trick. Providing monitoring and severe penalties for inadequate management could be fatal if the organization does not have sufficient management capability and the readiness not to deceive the public. Voluntary hostage posting should be used as an active device to raise trust with the backing of real competence and honesty towards the public.

Appendix A

The items used to measure prior attitude, estimation of trustworthiness, future incidents, and ending necessity of hostage in Experiment 1

Composite and items
Prior attitude
1. This company produces good products.
2. I like this company.
3. This is one of the top-ranked companies.
4. This company always does their best.
5. This is a good company.
6. I would like to buy their products.
Trustworthiness
7. This company is trustworthy.
8. This company deeply regrets the incident.
9. This company is honest.
10. This company has conscience.
11. This company cares about consumers.
12. This company can make full use of information to manufacture good products.
13. This company can get over its faults.
14. This company has a high level in technique of manufacturing.
15. This company has plenty of expert knowledge.
16. This company is capable of creating superior products.
Future incidents at the company
17. This company will not make a faulty product anymore.
18. This company will cause similar trouble again.
19. This company will continue to produce safe products in the future.
20. The products of this company will cause further incident.
21. The safety level of this company's products will decrease.
22. Somebody will get hurt by this company's dangerous product.
Ending necessity of a hostage
23. It is not necessary to keep watching this company.
24. This company should be obligated to allow free access to their information.
25. This company should be kept under close surveillance.
26. A boycott should be prepared on the assumption of deception by this company.
27. A prosecution and penalty system should be prepared for dishonest practices of this company.
28. A reward should be prepared for those who report dishonest practices of this company.

Appendix B

The items used to measure perceived integrity, benevolence, and participants' intention of usage of the company in Experiment 2

Composite and items
Motivation (Integrity)
1. This company is trustworthy.
2. This company is honest.
3. This company keeps promise.
4. This company is irresponsible.[a]
5. This company is reliable.
Motivation (Benevolence)
6. This company works for consumers' interests.
7. This company concerns about consumers' safety.
8. This company is concerned about consumers' thoughts.
9. This company has no perspective from consumers' point of view.[a]
10. This company makes efforts towards consumers' gratification.
Intention of usage
11. I do not want to buy their product anymore.[a]
12. Even if some of my family bought their product, I would not eat it.[a]
13. I would not like to go to this company's restaurant.[a]
14. I still would like to eat there even after reading about their incident.
15. I would not go to their restaurant even if my friend invited me there.[a]

[a] Reversal items.

References

Barber, B. (1983). *The logic and limit of trust*. New Brunswick, NJ: Rutgers University Press.

Covello, V. T. (1992). Trust and credibility in risk communication. *Health and Environment Digest, 6*, 1–3.

Cummings, L. L., & Bromiley, P. (1996). The Organizational Trust Inventory (OTI): Development and Validation. In R. M. Krammer & T. Tyler (Eds.), *Trust in organizations* (pp. 68–89). Newbury Park, CA: Sage Publications.

Cvetkovich, G., & Lofstedt, R. E. (1999). Social trust in risk management. In G. Cvetkovich & R. E. Lofstedt (Eds.), *Social trust and the management of risk* (pp. 1–8). London: Earthscan Publications.

Cvetkovich, G., & Winter, P. L. (in press). The what, how, and when of social reliance and cooperative risk management. In M. Siegrist, T. C. Earle, H. Gutscher, R. E. Lofstedt (Eds.), *Trust and risk management*. London: Earthscan Publications.

Cvetkovich, G., Siegrist, M., Murray, R., & Tragesser, S. (2002). New information and social trust: Asymmetry and perseverance of attributions about hazard managers. *Risk Analysis, 22*, 359–367.

Earle, T. C., & Cvetkovich, G. (1995). *Social trust: Toward a cosmopolitan society*. Westport, CT: Praeger Press.

Fischhoff, B. (1999). If trust is so good, why isn't there more of it. In G. Cvetkovich & R. E. Lofstedt (Eds.), *Social trust and the management of risk (Forward 3–5)*. London: Earthscan Publications.

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. Glencoe, IL: Free Press.

Garske, J. P. (1976). Personality and generalized expectations for interpersonal trust. *Psychological report, 39*, 649–650.

Gautschi, T. (1999). *A hostage trust game with incomplete information and fairness considerations of the trustee*. (ISCORE Paper No. 131). Utrecht, Netherlands: Utrecht University, Department of Sociology.

Gurtman, M. B., & Lion, C. (1982). Interpersonal trust and perceptual vigilance for trustworthiness descriptors. *Journal of Research in Personality, 16*, 108–117.

Keren, G., & Raub, W. (1993). Resolving social conflict through hostage posting: Theoretical and empirical considerations. *Journal of Experimental Psychology: General, 122*, 429–448.

Koren, G., & Klein, N. (1991). Bias against negative studies in newspaper reports of medical research. *Journal of the American Medical Association, 266*, 1824–1826.

Kraus, N., Malmfors, T., & Slovic, P. (1992). Intuitive toxicology: Expert and lay judgment of chemical risks. *Risk Analysis, 12*, 215–232.

March, J., & Olson, J. (1989). *Rediscovering institutions: The organizational bias of politics*. New York: Free Press.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*, 709–734.

Mayer, R. C., & Davis, J. H. (1999). The effects of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology, 84*, 123–136.

Metley, D. (1999). Institutional trust and confidence: A journey into a conceptual quagmire. In G. Cvetkovich & R. E. Lofstedt (Eds.), *Social trust and the management of risk* (pp. 100–116). London: Earthscan Publications.

Mlicki, P. (1996). *Hostage posting as a commitment device in the prisoner's dilemma game & hostage posting as a mechanism for co-operation in the prisoner's dilemma game. Codebook of two experiments on hostage posting*. (ISCORE Paper No. 102). Utrecht, Netherlands: Utrecht University, Department of Sociology.

Nakayachi, K., & Ohnuma, S. (2003). Trust and consensus building in the context of environmental risk management: A questionnaire-based research in Sapporo on the Chitose drainage canal plan. *The Japanese Journal of Experimental Social Psychology, 42*, 187–200 (in Japanese)..

Peters, R. G., Covello, V. T., & McCallum, D. B. (1997). The determinants of trust and credibility in environmental risk communication: An empirical study. *Risk Analysis, 17*, 43–54.

Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton, NJ: Princeton University Press.

Raub, W., & Keren, G. (1993). Hostage as a commitment device: A game-theoretic model and an empirical test of some scenarios. *Journal of Economic Behavior and Organization, 21*, 43–67.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*, 393–404.

Schweizter, M. E., Hershey, J. C., & Bradlow, E. T. (2003). Promises and lies: Restoring violated trust. *Paper presented at the 2003 annual meeting of academy of management*, Seattle, WA.

Shelling, T. C. (1960). *The strategy of conflict*. London: Oxford University Press.

Siegrist, M. (2000). The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. *Risk Analysis, 20*, 195–203.

Siegrist, M., & Cvetkovich, G. (2000). Perception of hazards: The role of social trust and knowledge. *Risk Analysis, 20*, 713–719.

Slovic, P. (1993). Perceived risk, trust, and democracy. *Risk Analysis, 13*, 675–682.

Slovic, P. (1999). Trust, emotion, sex, politics and science: Surveying the risk-assessment battlefield. *Risk Analysis, 19*, 689–701.

Slovic, P., Flynn, J., Johnson, S. M., & Mertz, C. K. (1993). *The dynamics of trust in situations of risk* (Report No. 93-2). Eugene, OR: Decision Research.

Williamson, O. E. (1983). Credible commitment: Using hostage to support exchange. *The American Economic Review, 73*, 519–540.

Yamagishi, T. (1998). *Shinrai no kouzou [Structure of trust*. Tokyo: University of Tokyo Press.

Yamagishi, T., Kikuchi, M., & Kosugi, M. (1999). Trust, gullibility, and social intelligence. *Asian Journal of Social Psychology, 2*, 145–161.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United state and Japan. *Motivation and Emotion, 18*, 129–166.