European Economic Review 137 (2021) 103811

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Escalating commitment to a failing course of action — A re-examination

Nikolai Martens, Henrik Orzen [*]

Department of Economics, University of Mannheim, L7, 3-5 Mannheim, 68131, Germany

ARTICLE INFO

JEL Classification Numbers:
C91
D81
D91

Keywords:
Experiments
Sunk-cost fallacy
Prospect theory
Reflection effect

ABSTRACT

We report the results from an experiment inspired by a seminal study (Staw, 1976) that introduced the notion of "escalation of commitment", a variant of the sunk cost fallacy. This topic has received much attention outside of economics, and we investigate whether the escalation phenomenon can be reproduced using standard protocols established in experimental economics. The focus is on how decision makers respond to a signal about a previous investment depending on (i) whether or not they were responsible for that investment, (ii) whether the signal is positive or negative, and (iii) whether or not the signal is associated with a loss or gain. We characterize theoretical conditions under which escalation—increased follow-up investments after receiving a negative signal—may occur. Our data indicates that subjects react differently to negative feedback when this feedback is linked to a financial loss and when they have been responsible for the initial investment. We also observe gender effects.

1. Introduction

Over the last decades a substantial body of work on the phenomenon of escalation of commitment to a failing course of action has emerged. The topic has received particular attention in the management and organizational behavior literature but has also been discussed in related fields (for surveys see Staw and Ross, 1987, 1989, and Sleesman et al., 2012, 2018). Escalating commitment is a manifestation of the sunk cost fallacy. Imagine a situation in which a decision maker realizes that a previously taken, costly course of action delivers disappointing intermediate results and then chooses to commit additional resources to the same course of action in the hope that outcomes will improve. The decision maker's rationale is that withdrawing from the originally chosen path at this point implies that any previous expenditure on it must be classified as wasted whereas some perseverance may lead to a result in which previous expenditures can be declared an investment that has ultimately paid off. In the typical scenario it is suggested that the bad news the decision maker receives contains information about future performance (the chosen path is "failing"). Thus, staying on course or getting even more invested is in fact foolish but the decision maker feels "locked in" and is therefore willing to throw good money after bad instead of cutting her losses.

Although there is a large amount of anecdotal evidence and a sizeable empirical literature, it is difficult to diagnose irrational escalating commitment conclusively and with precision in naturally-occurring settings because an observer has to reconstruct the parameters of the situation rather carefully in order to be able to judge whether, given the information available at the time, it was really ill-advised to become further engaged in the chosen path (Camerer and Weber, 1999). A second problem is that decision makers

* Corresponding author.
E-mail address: henrik.orzen@uni-mannheim.de (H. Orzen).

https://doi.org/10.1016/j.euroecorev.2021.103811
Received 8 September 2020; Received in revised form 26 April 2021; Accepted 11 June 2021
Available online 21 June 2021
0014-2921/

who act not exclusively on their own accord but at least partially on behalf of someone else or a collective may have individual incentives to hide that wrong decisions have previously been made (Leeson, 1996). Finally, there may be strategic, reputational motives for being seen to stick to a chosen path (Kanodia et al., 1989; Berg et al., 2009).

Alternative rationalizations for seemingly escalating behavior can be largely ruled out in suitably designed laboratory settings where the experimenter has the power to control what information is made available to participants, which actions they can take and what the incentives are. Indeed, the seminal paper that marked the starting point for the escalation literature reported the results of an experiment (Staw, 1976) and was followed by many other experimental studies across a number of (sub-)disciplines. However, the topic—or indeed the sunk cost fallacy more broadly—has not received much attention in the field of experimental economics. This is surprising for at least two reasons. First, the phenomenon of escalation of commitment seems highly relevant for economics as it constitutes another challenge for the standard *homo economicus* paradigm (Thaler, 1980) and has been associated with a number of high-stake economic disasters (see Sleesman et al., 2012, for a list of examples). Second, while there is a large experimental literature in psychology and management, there are distinct differences in what are deemed acceptable experimental procedures and protocols between these disciplines and experimental economics (Hertwig and Ortmann, 2001; Croson, 2005). In particular, it is noteworthy that the evidence from sunk cost fallacy experiments in which the decisions of subjects have real monetary consequences for them is mixed (Phillips et al., 1991; Friedman et al., 2007; Buchheit and Feltovich, 2011; and Haita-Falah, 2017).[1]

In this paper we revisit Staw's (1976) seminal work that established the notion of escalation of commitment as an important subject of investigation and a potential driver of biased economic decision-making. The participants in Staw's experiment were asked to imagine themselves in the role of a corporate executive who allocates research and development funds between the two main divisions of a fictional company. Treatments varied in terms of feedback given to subjects and in terms of the responsibility for an initial allocation decision. The remarkable main finding was that subjects increased their investments on their initially chosen division when they received information that it was *failing*. We design a new laboratory experiment in the spirit of this original study but employ standard methods of modern experimental economics. In particular, we introduce monetary incentives, take out the rich background story, control for self-selection, provide unequivocal information and feedback to subjects, and isolate the effects of responsibility more stringently. Our project is motivated by the fact that, as we will discuss in more detail below, the behavior that Staw observed in his experiment leaves room for interpretation. The goal is to either provide clearer support for the phenomenon and ascertain its relevance for economics, or to uncover weaknesses in the existing evidence and raise concerns about efforts to explain real-world economic decisions by ascribing them to irrational escalation of commitment to a failing course of action.

Is it of interest to reexamine a more than forty-year-old paper? We think it is, for a number of reasons. First, Staw (1976) has had an enormous and lasting impact and is regularly cited to this day but to the best of our knowledge there has been no attempt to test whether his findings are replicated when payoffs are not merely hypothetical and more conventional protocols are used. Second, the specific kind of setting that Staw investigates bears some resemblance to real-world scenarios in which the presence of escalating commitment may have particularly undesirable consequences (e.g. financial decision-making, project management, or bank loans). Third, while there are now a few studies in experimental economics on the sunk-cost effect more broadly, we believe that none of them focuses on how subjects respond to positive versus negative signals about the chosen course of action or on the role of personal responsibility.[2]

In our setting subjects are given the opportunity to invest in lotteries with two possible outcomes, "success" or "failure". There are two stages. In the first stage subjects make a binary choice between purchasing a lottery and keeping their endowment. In the second stage subjects make a near-continuous investment choice such that they can decide themselves what fraction of their endowment to keep and how much to spend on the lottery (as in Gneezy and Potters, 1997). Subjects know that the probability of success is the realization of a single random draw from a uniform distribution on the interval $(0, 1]$ and that the probability of success is the same in both stages but they do not know the probability itself. The outcome of the first stage therefore provides a signal about whether or not the "lottery project" is failing or succeeding.

Our findings are as follows. First, in one of our control treatments—in which subjects are not responsible for the stage-1 decision and do not experience a gain or loss from the stage-1 outcome—we observe an intuitively appealing response pattern: A negative signal about the probability of success causes a systematic reduction in investments in stage 2 although this response is less strong than one would expect theoretically. Second, in our main treatment—in which subjects are both responsible for stage 1 and suffer or enjoy the consequences—we find that average investments are unaffected by the signal. This is specifically due to a change in the response to a *negative* outcome relative to the control treatment, indicating escalation of commitment. However, unlike Staw we do not find that average investments are actually boosted by negative feedback. Third, to disentangle the effects of experiencing a loss or gain and the effects of responsibility we run a second control treatment in which responsibility for the stage-1 choice is removed but the financial consequences are not. Our results suggest that both factors contribute to the differences between the main treatment and the first control treatment. Fourth, there are quite strong gender effects. Men are not only less risk averse than women but they are also considerably more responsive to the signal in all treatments.

Thus, although we cannot confirm the extreme reaction to bad news that Staw observed, we do find that subjects respond

[1] There is evidence that loss avoidance can influence equilibrium selection in coordination games (Cachon and Camerer, 1996; Rydval and Ortmann, 2005; Feltovich et al., 2012) but a major factor appears to be that players mainly expect *others* to be affected by the desire to avoid losses.

[2] Weber and Zuchel (2005) investigate how subjects' risk attitudes are affected by prior wins versus losses from investments that they have either chosen themselves or that have been assigned to them. However, in their paper the intermediate feedback contains no information on the future profitability of the investment.

differently to a negative signal depending on whether they are responsible for the first decision and depending on whether the negative signal comes with a monetary loss. Furthermore, similar to Staw we find no significant difference in the response to a *positive* signal across treatments.

2. Staw's experiment

2.1. Summary

Staw reports that he recruited his subjects from an undergraduate business course. Participation in the experiment was one way of fulfilling a particular course requirement. Subjects were asked to work on a hypothetical case scenario in which they assumed the role of a corporate executive whose job it is to allocate funds for research and development within a fictional company.

The first task in the *high personal responsibility* treatment was to make a binary choice of investing a fixed amount of 10 million dollars in either the 'Consumer Products' division or the 'Industrial Products' division. As a basis for their decision subjects were shown historical data on sales and earnings indicating that the profitability of both divisions had deteriorated substantially in recent years, and the background story provided additional contextual information. The paper contains no details on this additional information but presumably it was rich and ambiguous enough so that it was possible to make a judgment call favoring either option.[3] Subjects had to justify their decision in a brief written statement and were required to write their names on their answer sheets.

After completing stage 1 subjects received new material that provided information on how the two divisions had performed in the five years following the initial binary decision. Half of the subjects got the feedback that their chosen division had improved while the non-chosen division had further declined, and half of the subjects got the exact opposite feedback. The task in stage 2 was to allocate further funds, a total of 20 million dollars, between the two divisions. However, subjects were now allowed to split up the money in any way they liked. Again, they were asked to defend their decision in writing and non-anonymously.

For the *low personal responsibility treatment* the exact same case scenario was used but this time the stage-1 decision became part of the background story (the binary R&D funds decision "had been made [...] by another financial officer of the company", p.32) and subjects were only responsible for the stage-2 allocation of the 20 million dollars. Within this treatment there were four information conditions, evenly distributed among the participants, that varied which division had been selected by the other officer and which division had performed well and which one badly following the stage-1 allocation decision.

Staw found that positive or negative feedback about the earlier chosen division made little difference to the stage-2 investments in the low responsibility treatment, and investment allocations following positive feedback in the high responsibility treatment were again similar. However, *negative* feedback about the alternative that subjects had previously selected *themselves* led to a substantial and significant increase in the money allocated to the same alternative. This was the main result and it was taken as evidence that people tend to irrationally escalate their commitment to a previously chosen but now failing course of action. While we agree that this a possible interpretation there are a number of uncertainties. This is due to some details of the experimental design, which we discuss in the following. This discussion also motivates several of our own design choices.

2.2. Discussion

Staw's study provides guidance on what escalation of commitment really means and why it may be relevant. While his experimental design is also ingenious in many ways, we nevertheless see several problems that raise questions about what we can infer from the data. In particular, the rich background story gives a lot of space for idiosyncratic and diverging interpretations of the setting.

First, there are no well-defined criteria for what constitutes a "good" decision or the firm's goals. This could be profitability within the time window that is shown in the case material. But it could also be long-term presence of the firm in the two market segments or helping the division that is in most trouble. What adds to this is that the return from R&D investments are known to potentially take a particularly long time.

Second, it is not clear what to make of the feedback provided. The interpretation favored in the paper is that negative feedback about a division's performance signals that this division is on an inevitable path of failure, which makes investing in it unwise. This is key if we want to arrive at the conclusion that the observed behavior is to be judged as 'biased'. However, this reading is not at all obvious from a participant's vantage point. For example, participants may believe that the early similarly poor trajectories of the two divisions or the later divergent developments cannot be expected to continue in the same fashion. Thus, rather than extrapolating the provided financial data linearly they may interpret a certain dip in a division's performance as an indication that now is a particularly good time to invest, effectively betting on a kind of boom-and-bust cycle.

Third, subjects receive no information on how their decisions will affect the fortune of the firm. Prior to their initial choice, participants are led to believe that insufficient R&D expenditures have been the primary reason for the firm's poor performance in the first place. If a decision-maker now learns that the previously supported division continues to do badly, it may be quite reasonable for him or her to assume that the initial 10 million dollars have just not been enough to turn the ship around. Perhaps the earlier commitment on the "failing" division has prevented outcomes that would have been even worse, which would then justify additional investments. In contrast, the more successful division may be seen as requiring less attention. The general problem here is that subjects

[3] The data shows that about 54% of subjects ended up choosing the consumer products division while the remainder preferred the industrial products division.

have to make their stage-2 decision without knowing what the counterfactual outcome from the stage-1 decision would have been. This is realistic, of course, but it provides an obstacle for drawing conclusions from the data because it could be that subjects may interpret the situation in different ways.

Fourth, a popular explanation for Staw's results is the notion of "self-justification" according to which people have the wish to maintain a self-image of consistency and find it easier to rationalize an earlier decision that has led to adverse consequences than to suffer from cognitive dissonance (Festinger, 1957) caused by the admission that they have made a mistake. It can be argued, however, that the required written explanations trigger more than a desire for mere *self*-justification. Given the context in which the experiment took place subjects may have seriously considered the possibility that they would be in some form assessed based on their written statements. This may have caused a sense of external pressure for consistency.

Fifth, Staw does not control for self-selection effects (see also Schulz-Hardt et al., 2009). Subjects in the low personal responsibility treatment are assigned at random to a history in which a previous decision maker has opted for the Consumer Product division or to the alternative history in which the Industrial Product division was chosen. In contrast, subjects in the high personal responsibility treatment who—based on their personal interpretation of the setting—are more inclined to see potential in, say, the Consumer Product division will tend to favor this division in both stages.

Finally, an unsurprising comment from the viewpoint of experimental economics is that Staw's design does not feature actual monetary incentives. It is usually argued that without such incentives, subjects have little reason to take the task at hand seriously. The mere request "to do the best job they could on the cases" and "to make a good financial decision" (Staw, p.31) may create insufficient motivation to consider the options carefully and to pause and reflect on a first, impulsive intuition of what to do. However, it should be noted that Staw's student participants fulfilled a course requirement by taking part in the experiment and, being business students, may have seen the task as a useful exercise in improving their skills for similar tasks in graded examinations.

3. Experimental design and hypotheses

We now turn to our own design which is more conventional from an experimental economics perspective. We begin with a description of the main two-stage task in our *responsibility* treatment RESP.

3.1. The RESP treatment

At the beginning of stage 1 of this treatment subjects were endowed with 150 tokens ("Talers") and then faced a binary choice between (i) keeping them all and (ii) spending all of them on a lottery that would yield either 300 tokens ("success") or 75 tokens ("failure") with a 50–50 chance.[4] We explained to subjects that the outcome of the lottery would be determined, separately for each individual, via two random draws (with replacement) from a bag of 100 consecutively numbered plastic chips. The first draw would determine a subject's "personal success probability"—an integer number between 1 and 100 corresponding to a probability from the set $\{0.01, 0.02, \ldots, 1.00\}$—and the second draw would determine the actual outcome by comparing the number on the drawn chip to the personal success probability number. If the second draw produced a number less than or equal to the first number, the subject would obtain 300 tokens; otherwise (s)he would obtain 75 tokens.[5]

All random draws were physically conducted and recorded in the laboratory's control room adjacent to the lab itself. This was done with the help of one of the subjects who had been randomly selected at the beginning of the session out of all participants present. The role of this assistant was to lend a hand in making the draws and double-checking the recordings but more importantly to certify, as a representative of our participants, that all procedures were carried out exactly as announced. The assistant was paid a lump-sum of 8 Euros.

As soon as the draws were completed each subject was informed about the outcome of his or her stage-1 lottery ("success" or "failure"). Note that we ran the lottery and informed subjects about its outcome irrespective of their choice in stage 1. Thus, even those who had decided to keep their endowment in stage 1 found out what would have happened if they had played the lottery. Note also that we did not reveal the actual numbers drawn until the very end of the session. In particular, subjects were not yet told the value of their personal success probability.

At the beginning of stage 2 subjects received an additional 300 tokens each. Their task now was to decide how many of these to keep and how many to invest in another lottery, knowing that the outcome of this second lottery would be determined by another single draw from the bag of plastic chips.[6] If the number of the drawn chip was less than or equal to the personal success probability number from stage 1, the outcome would be "success" and the number of tokens invested in the lottery would be doubled; if not, the outcome would be "failure" and the number of tokens invested in the lottery would be halved.

Some aspects of our RESP treatment closely resemble Staw's *high personal responsibility* treatment. Subjects first make a binary investment choice, they then receive feedback that contains either positive or negative news about their chosen alternative, and finally they make a (near-)continuous choice of dividing up a sum of money which is twice as large as the amount from the first stage. Other

[4] See Appendix A for the experimental instructions.

[5] The exact probability of success was 0.505 (see Appendix B for details). Had we included a "0" chip for the first draw it would have been exactly 0.5 but we chose to eliminate the possibility of a personal success probability of zero.

[6] Specifically, subjects could allocate any whole number of tokens $x \in \{0, 1, \ldots, 300\}$ to the lottery option knowing that the remainder, $300 - x$, would be automatically allocated to the safe option.

aspects depart from Staw's setting. One of the principles guiding our design choices was that we wanted subjects to face a context-free, yet meaningful personal decision problem in each stage. We drop the requirement to explicitly justify one's choices because our focus is on the question whether escalation of commitment occurs in a free-choice setting without external pressures. In our experiment subjects do not act as managers representing a company but make decisions on their own behalf. There is no background story. Subjects' decisions have direct monetary consequences for them. Our two options evidently differ in terms of expected value and in terms of risk, whereas in Staw's design both options were characterized by uncertainty. The feedback in between the two stages is not assigned by the experimenter but is determined by the outcome of a lottery and subjects know this. The feedback refers to only one of the options (the lottery) since the outcome from the alternative path (the safe option) is clear from the start. This also implies that regardless of a decision maker's stage-1 choice she will know the counterfactual outcome when she makes her stage-2 choice. Finally, the feedback contains a clear signal about the personal success probability, i.e. about how promising the initially chosen course of action is relative to its alternative for a follow-up investment.

While we will discuss theoretical benchmarks and derive predictions from them in the following subsection, there is a first behavioral hypothesis we can formulate already:

Hypothesis 1. Gender. Women invest less in the continuous token allocation task than men and are less likely to take the lottery in the binary choice task than men.

This hypothesis suggests itself due to the fact that our stage-2 task corresponds to the investment task introduced by Gneezy and Potters (1997) and it is often reported that women invest less than men in this task. Charness and Gneezy (2012) collect data from several studies and find strong evidence for the robustness of this finding. We hypothesize that women will also be less inclined to pick the lottery in the binary choice task.

3.2. Theoretical considerations

3.2.1. Expected utility maximization

The initial lottery option in the *RESP* treatment is a relatively good bet with an expected value of 188.6 tokens. Consider the exponential utility function

$$u_i(y) = \begin{cases} \dfrac{1 - e^{-\alpha_i y}}{\alpha_i} & \text{if } \alpha_i \neq 0 \\ y & \text{if } \alpha_i = 0 \end{cases} \tag{1}$$

where $y > 0$ denotes individual i's payoff and α_i is i's risk preference parameter (with $\alpha_i > 0$ indicating risk aversion).[7] The individual will decline the first lottery if $\alpha_i > 0.0066$. A person with a risk preference parameter close to this threshold level could be described as moderately risk-averse: At $\alpha_i = 0.0066$ the certainty equivalent to a gamble that pays \$100 or nothing with a 50–50 chance is about \$42.

Since the personal success probability remains the same in stage 2, the outcome of the first lottery contains information about whether the initially taken course of action is failing or succeeding. In fact, the roughly even chance of a positive outcome in stage 1 increases to a posterior probability of about 2/3 or drops to a value marginally above 1/3 after the first outcome is revealed.[8] Under expected utility theory this signal affects the optimal stage-2 investment in an intuitive way: A positive signal encourages investments while a negative one acts as a deterrent. This is illustrated in Fig. 1.

Furthermore, as can also be seen in Fig. 1, if the first lottery leads to a successful outcome then anyone who had previously found it optimal to choose this lottery ($\alpha_i \leq 0.0066$) should now feel encouraged to also invest a significant proportion of her stage-2 endowment, or even all of it, in the second lottery.[9] But even someone who found the first lottery too risky may allocate a moderate number of tokens to the lottery unless she is extremely risk-averse.

On the other hand, if the signal is bad ("failure"), we would expect that anyone who had initially chosen the safe alternative should now feel confirmed in her decision (and invest next to nothing, see Fig. 1 again), whereas someone who had opted for the lottery should now question their attachment to the lottery option. Note, however, that the expected profit from investing under the posterior probability of success remains positive (it is close to zero though: the expected return from 100 invested tokens is 100.5 tokens). This is a deliberate aspect of our design, as we want to study a setting in which there is some scope for individuals who may be susceptible to escalation of commitment to rationalize investing in the second lottery. A too strong signal from a stage-1 "failure"—such that any

[7] This function has some appealing properties that are useful in particular when we later extend our model to include loss aversion and different risk attitudes in losses versus gains (see Köbberling and Wakker, 2005).

[8] Appendix B provides details on our theoretical results.

[9] The optimal investment at the threshold value of $\alpha_i = 0.0066$ is coincidentally very close to 150 tokens (the precise value is 142). Thus, assuming that the average person choosing the lottery in stage 1 is not extremely close to indifference we should expect that those who prefer the lottery option invest on average more tokens in stage 2 than in stage 1.

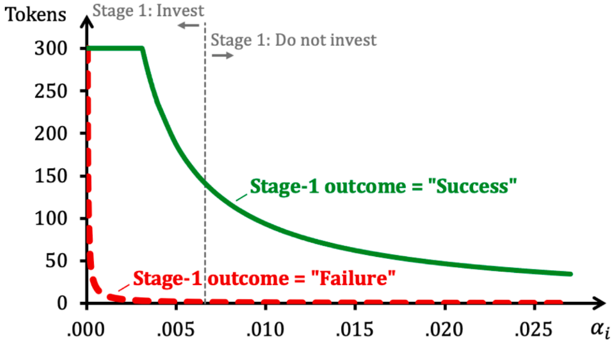

Fig. 1. Optimal stage-2 investment as a function of risk aversion and feedback.

further investment is clearly a ridiculously bad idea—would run the risk of drowning out any tendencies towards the bias and would not provide a very informative test.[10] Of course, this leaves the possibility that some risk-neutral individuals simply invest all their tokens in both stages, without this indicating any form of sunk-cost fallacy. Our additional treatments will help us to control for this, as will become apparent below.

As we expect our subjects to display some degree of risk aversion on average, we formulate.

Hypothesis 2. RESP EUT. (a) If the stage-1 lottery outcome is "success" then subjects who chose to invest their 150 tokens in the stage-1 lottery invest even more in the stage-2 lottery. Subjects who chose the safe option in stage 1 invest more moderately. (b) If the feedback from the first lottery is "failure" investments drop to minimal levels.

3.2.2. *Prospect theory*

Prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) has been discussed as an explanation for escalation of commitment at least since Whyte (1986). What does it predict in our setting? Consider a simple extension of our earlier model [1] to incorporate reference dependence and loss aversion:

$$u_i(y) = \begin{cases} \dfrac{1 - e^{-\alpha_i y}}{\alpha_i} & \text{if } y \geq 0 \\ -\lambda_i \dfrac{1 - e^{\alpha_i y}}{\alpha_i} & \text{if } y < 0. \end{cases} \tag{2}$$

In this model y is the absolute gain or loss that individual i experiences relative to a given reference point, $\lambda_i \geq 1$ is the individual's loss-aversion parameter and $\alpha_i \geq 0$ describes the individual's risk preferences for gains and losses.[11] As before, when $\alpha_i = 0$ the fractions in the expressions are to be replaced by y.

Taking the initial endowment of 150 tokens as the reference point and assuming $\alpha_i = 0$ for now, the expected payoff from playing the lottery is $150p_1 - 75\lambda_i(1 - p_1)$ while keeping the endowment yields 0. With a probability of success of $p_1 = 0.505$ the decision maker is predicted to choose the safe option if $\lambda_i > 2.04$. If α_i is positive the required degree of loss aversion is lower. For more details see Fig. 2 and Appendix B.

The decision maker's preferred actions in both stages not only depend on the α and λ parameters but are also affected by the reference point against which she evaluates her realized earnings. In principle, the reference point could be formed in various ways and at various points in time. For example, subjects may form certain expectations about the payment they receive in an experiment prior to taking part, and then assess their actual earnings against this benchmark. Alternatively, they may perceive *any* payment in an experiment as a gain. Or they base their reference point on the initial endowment of 150 tokens and then compare any further developments in the entire experiment to this. There are many possibilities.

We favor the idea that the reference point in each stage is given by the total endowment received so far.[12] Thus, we assume that a participant who evaluates her post-stage-2 payoffs against a reference point will reason as follows: "The experimenters have endowed me with 450 tokens in total and they have given me the option to just keep these and do nothing. Alternatively, I had the opportunity to make risky investments in stage 1 and stage 2 that I knew may lead to additional earnings or lead to losses." Under this assumption the

[10] For similar reasons we chose not to calculate posterior probabilities and expected returns for our subjects. Decision makers in the real-world examples that are discussed in the literature do not have access to exact probabilities and expected values either, and some ambiguity about the situation may be an important prerequisite for escalation of commitment. Thus, we wanted to maintain a sense of uncertainty about how profitable the lottery option is.

[11] Note that we employ here a simplified prospect-theoretical model without probability weighting and with only a single curvature parameter for the utility function. Function [2] does, however, implement prospect theory's reflection effect: The decision maker is risk-averse in the domain of gains ($y \geq 0$) but risk-prone in the domain of losses ($y < 0$).

[12] In stage 1 subjects had no knowledge yet that they would receive an additional 300-token endowment at the beginning of stage 2.

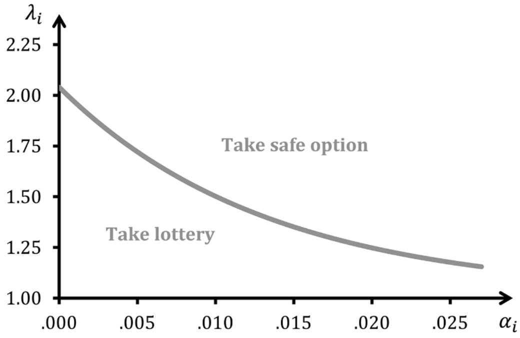

Fig. 2. Stage-1 decision as a function of α_i and λ_i parameter values.

decision maker will find that her point of departure at the time she chooses her second investment is, dependent on her stage-1 choice and on the outcome of the lottery, either (i) $D_2 = 0$ (she has chosen the safe option in stage 1); (ii) $D_2 = 150$ (she has chosen and won the stage-1 lottery and is now in the gains domain); or (iii) $D_2 = -75$ (she has lost the stage-1 lottery and is now in the loss domain).

Table 1 lists predictions from our model for different values for α_i, λ_i and D_2. Consider first the case of a positive signal such that an individual who played the first lottery now finds herself in the domain of gains ($D_2 = 150$). As the table shows, the model predicts that such an individual would in this situation allocate a substantial part of her tokens to the second lottery provided that she is not too risk averse. Loss aversion plays no role because even if she risks all her 300 tokens the worst that could happen is that she moves back to her reference point.

The comparison with the case of a negative signal where the individual's position is in the loss domain ($D_2 = -75$) delivers ambiguous results. If the individual is not loss averse ($\lambda_i = 1$) she is predicted to allocate her entire stage-2 endowment to the lottery regardless of the value of α_i. This is in stark contrast to our earlier results in the model of expected utility maximization and is due to prospect theory's reflection effect: Accruing losses due to a bad stage-2 lottery outcome—worsening the decision maker's position further—does not impact proportionally on the utility scale because her utility function is convex in the domain of losses. As a result, she is prepared to enter a high-stakes bet.

Once we allow for loss aversion, her willingness to invest may be reduced substantially (see Table 1 again). Yet, even when loss aversion is important, investing 75 tokens dominates any lower investment and the optimal investment can rise again to much higher levels when α_i is increased, again due to risk proneness in the loss domain. However, there is a selection effect: Decision makers with a high loss aversion parameter *and* a high α_i would not have chosen the first lottery. That is, they cannot now *be* in the position $D_2 = -75$. Fig. 3 provides examples to illustrate this.

In summary, without specific information on an individual's attitude towards losses versus gains and on her risk attitudes prospect theory does not deliver clear-cut comparative-static predictions for investment decisions following success or failure in playing lottery 1. However, there are parameter constellations for which the model predicts that an individual invests as many, or even more, tokens after having received a negative signal than after a positive signal, the pattern that Staw observed in his experiment.

Finally, consider a scenario in which the individual opted against the stage-1 lottery ($D_2 = 0$). For the case of risk neutrality ($\alpha_i = 0$) the stage-2 prediction is now either full investment (all 300 tokens) or no investment at all (see Table 1 once more). This depends on the loss aversion parameter. For a negative signal the individual invests only if $\lambda_i < 1.015$, i.e. she must be virtually loss neutral for risking her tokens to make sense. The corresponding condition for a positive signal is $\lambda_i < 4.061$. Hence, in this case the individual should hold back only if she is *extremely* loss averse. Increasing the α_i parameter reduces the investment activity, as shown in Table 1. Fig. 4 shows two more examples.

The predictions for the $D_2 = 0$ case are qualitatively similar to the results from expected utility theory (Fig. 1). Overall, the prospect theory model allows the possibility of escalation of commitment and a contrarian signal response, at least for some parameter constellations. For an illustration, consider an individual with $\lambda_i = 1.25$ (left panel in Fig. 3 and Fig. 4) who is indifferent between taking the lottery or the safe option in stage 1 ($\alpha_i = 0.0199$). If that individual went for the safe alternative, a negative signal would cause her to avoid investing in stage 2 (Fig. 4), just like a similarly risk-averse expected utility maximizer. But if the same individual received the negative signal after having selected the lottery, she would subsequently invest almost all her stage-2 tokens (Fig. 3) and thus display escalating commitment to her chosen path. Furthermore, her investment would be substantially larger than if the signal had been positive.

Hypothesis 3. RESP PT. (a) If the stage-1 lottery outcome is "success", subjects invest moderately in the stage-2 lottery. (b) If the feedback from the first lottery is "failure" then those who had preferred the safe option in stage 1 hardly invest at all in stage 2. (c) However, those who had taken the risky option and then experienced "failure" invest substantial amounts, at least as much as their counterparts who got the feedback "success".

Table 1
Predictions for stage-2 investments.

D_2	λ_i	α_i 0.000	0.005	0.010	0.015	0.020	0.025
Following the event "success"							
150	any	300	187	93	62	47	37
0	1.0	300	300	280	187	140	112
	1.5	300	300	199	133	100	80
	2.0	300	283	142	94	71	57
	3.0	300	121	61	40	30	24
	4.0	300	6	3	2	2	1
	≥ 4.1	0	0	0	0	0	0
Following the event "failure"							
0	< 1.1	300	6	3	2	1	1
	≥ 1.1	0	0	0	0	0	0
–75	1.0	300	300	300	300	300	300
	1.5	75	144	222	248	261	269
	2.0	75	75	164	210	232	246
	3.0	75	75	83	156	192	213
	4.0	75	75	75	117	163	190

Notes: D_2 indicates the point of departure, relative to the reference point, at the moment of the stage 2 investment decision. An individual who chose to play the first lottery would at this moment have an account balance of either 300 (stage 1 payoff) + 300 (new stage 2 endowment) = 600 tokens or 75 + 300 = 375 tokens. If that individual thinks of the total endowment from both stages as her reference point, then her point of departure is either $D_2 = 600 - 450 = 150$ or $D_2 = 375 - 450 = -75$. The account balance of an individual who had chosen *not* to play the first lottery would now be 450 tokens and thus $D_2 = 0$.

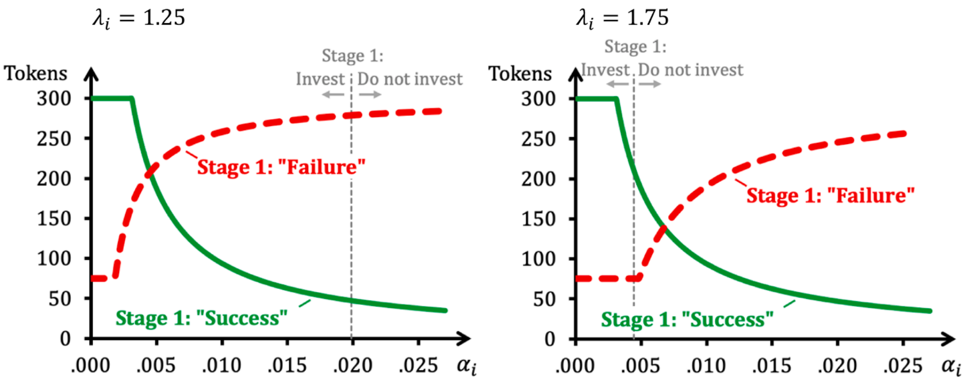

Fig. 3. Lottery-2 investments for $D_2 = -75$ and $D_2 = 150$.

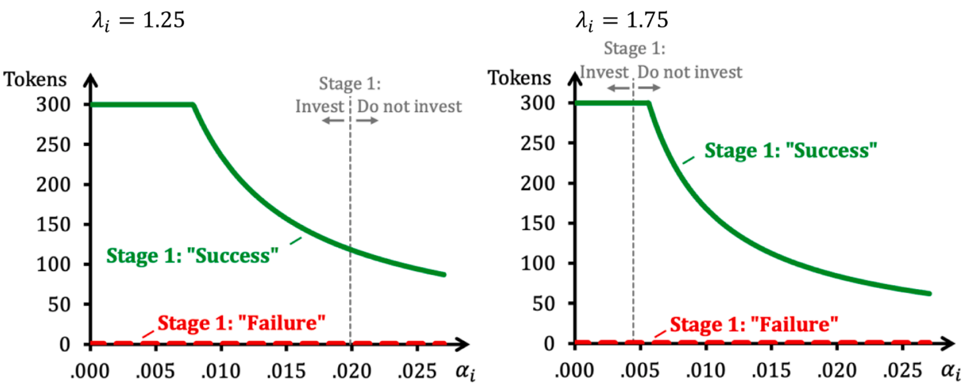

Fig. 4. Lottery-2 investments for $D_2 = 0$.

Note that our prospect-theoretical benchmark is not compatible with the idea that the safe option is perceived as a 'failing course of action' after a decision maker who took this option is informed that the lottery would have yielded the outcome "success". This is because we do not consider the lottery itself as a (stochastic) reference point.[13]

3.3. The RESUME treatment

In this treatment subjects were informed about the stage-1 task but were not asked to make the stage-1 decision. Instead, we linked each subject in the RESUME treatment to an anonymous "predecessor" who had taken part in the RESP treatment. Subjects were told what their predecessor had decided to do in stage 1 and what the outcome of the stage-1 lottery had been. However, they were not yet informed about the value of their predecessor's personal success probability. Subjects then received 300 tokens and had to decide how many of them to invest in the lottery *with the same personal success probability*, exactly in the same way as participants did in stage 2 of the RESP treatment.

This treatment is designed in the spirit of Staw's low personal responsibility treatment. Although subjects effectively take over from their predecessors and then resume the experiment in stage 2, they are neither responsible for the RESP stage-1 decision nor does the feedback (success or failure) cause them to experience any gain or loss. Instead, the past lottery outcome merely delivers a signal about the relevant probability of success.[14] Because our interest is focused on the effects of removing the gain/loss experience we collected data only for scenarios in which the (in other respects randomly selected) predecessor *had* chosen the first lottery.

From the vantage point of expected utility theory the differences between the RESP treatment and the RESUME treatment are immaterial: In both settings the only relevant factors for a decision maker's stage-2 investment choice should be her risk attitude and the nature of the signal. As shown before in Fig. 1, if the signal is negative—and she is at least marginally risk averse—she should invest very little but if the signal is positive it is optimal for her to invest most of her 300-token endowment.

However, Fig. 1 also shows that we should expect, particularly in the case of a positive signal, investments to be dependent on whether or not the lottery is preferred in the binary (stage-1) decision problem. To control for the individual's preference, we added a third stage to the RESUME treatment in which subjects faced exactly the same binary choice problem that their counterparts in the RESP treatment had been confronted with in stage 1. To reduce potential contaminations from the different experiences that subjects made prior to this choice, we did not reveal the outcomes from the stage-2 lottery until after they had made their stage-3 decision.

Hypothesis 4. RESUME EUT. Average investments in the stage-2 lottery of RESP and RESUME are the same. However, RESP participants who chose [did not choose] the stage-1 lottery invest more [less] in stage 2 than RESUME individuals do. Once we control for the decisions of RESUME participants in stage 3, these differences disappear.

From the vantage point of prospect theory, the RESUME treatment introduces an important difference: Unlike in RESP all subjects make the stage-2 decision at $D_2 = 0$. Thus, the predicted response to "success" versus "failure" is now more in line with expected utility theory (see Fig. 1 and Fig. 4) than in RESP. This leads to

Hypothesis 5. RESUME PT. Any tendencies towards escalation in RESP are not present in RESUME: Investments in response to "failure" are lower than in RESP and are lower than investments in response to "success".

3.4. The INHERIT treatment

This treatment is identical to the RESUME treatment except that subjects are paid not only for their own decisions but also inherit their predecessors' stage-1 earnings. The intention for running this treatment is to reintroduce a gain versus loss experience without returning the responsibility of the stage-1 decision to the decision maker. This treatment serves as a control for the 'responsibility effect' that Staw discussed in his paper. Since both expected utility theory and prospect theory are mute on possible effects of being responsible for a decision or not, neither of them predicts a difference between the INHERIT and the RESP treatment in terms of stage-2 investments.

Hypothesis 6. INHERIT EUT-PT. With and without controlling for the binary lottery-vs.-safe-option choice there is no difference in stage-2 investments between RESP and INHERIT.

However, if personal responsibility for the stage-1 decision *is* an important driver for any observed differences in stage-2 investments in RESUME relative to RESP then such differences should remain present in the INHERIT treatment. Self-justification theory, a leading explanation for Staw's results that is discussed in the escalation literature (see Staw, 1976; Whyte, 1986; Brockner, 1992), argues that being responsible for a negative outcome triggers a desire to retroactively rationalize the decision that led to that outcome in order to reduce cognitive dissonance. A consequence of this rationalization is that the decision maker becomes more committed to the initially chosen path than in a scenario without personal responsibility for the earlier decision. In contrast, no responsibility effect

[13] Of course, it could turn out that our choice of reference point is not correct. Escalation of commitment may then appear in the form of decision makers choosing the safe option in stage 1 but then, after having observed a positive signal, making extremely high investments in stage 2.

[14] Arguably, we could have simplified the instructions by removing the historical predecessor frame and presenting the random draw directly as a procedure to generate a signal about the unknown success probability. However, this would have changed the context to a greater extent than we were comfortable with.

is predicted for the case of positive feedback because there is no cognitive dissonance that needs to be resolved.

Hypothesis 7. Escalation due to self-justification. There are no treatment differences in stage-2 investments for subjects experiencing "success" but in response to a negative stage-1 outcome RESP subjects invest more than RESUME and INHERIT subjects.

Finally, we ask whether our results will replicate Staw's observation of soaring investments after negative feedback for the (self-) chosen path.

Hypothesis 8. Responsibility-based contrarian signal response. When RESP subjects who choose the stage-1 lottery experience "failure" they invest more in stage 2 than when they experience "success". This pattern is not replicated in RESUME and INHERIT.

An overview of our three treatments is provided in Table 2. We summarize the main predictions as follows. According to *expected utility theory* decision makers will base their investment in stage 2 on the signal about the profitability of the lottery option in all treatments. A negative signal implies minimal investments for any at least moderately risk averse agent. For the RESUME treatment *prospect theory* and the notion of *self-justification* are not at odds with expected utility theory. However, they predict upward-biased investments for the negative-signal case of the RESP treatment. Both also allow the possibility that such escalation of commitment may lead to a 'contrarian signal response' such that investments in the RESP treatment are higher when the signal is negative than when it is positive. Prospect theory attributes this bias to the experience of losses and thus sees no fundamental difference between the RESP and the INHERIT treatment. In contrast, for the self-justification approach personal responsibility is key and it therefore does not discriminate between RESUME and INHERIT.

3.5. Procedures

We ran the sessions as pen-and-paper experiments. Our motivation for doing so was as follows. First, we hoped that using a physical randomization device for determining the outcomes, especially the personal success probability, would be particularly credible and easy to grasp intuitively. Furthermore, inviting one of the subjects to inspect our procedures and to participate in determining the outcomes really only made sense with non-electronic random draws. Second, we thought that filling in a paper form would create a stronger sense of commitment than clicking buttons on a screen. Our procedure also created some natural delays which perhaps encouraged participants to consider their choices more carefully than with a speedy electronic interface that might have fostered a tendency to make hastier decisions. For the INHERIT treatment it was particularly convenient to have a natural delay in between the first general instructions that informed subjects about the 150-token endowment and the second set of instructions that revealed the outcome of the predecessor's stage-1 lottery. We think that this delay made it more likely that subjects perceived the initial 150 tokens as a reference point, giving prospect theory a better shot.

Subjects were recruited via ORSEE (Greiner, 2015). The number of participants was 138 (55 in RESP, 38 in RESUME and 45 in INHERIT), 82 of whom were males. Sessions lasted about 40–45 min, including the time for reading instructions and filling out questionnaires. Average earnings amounted to €8.50.[15]

4. Results

In this section we will make several between-subject comparisons relating to subjects' investments in the stage-2 task. The p-values we report in that context all refer to the results from a non-parametric two-sided two-sample Fisher-Pitman permutation test, unless explicitly stated otherwise.

4.1. A first look at binary choices and lottery investments

As shown in Table 3, about four out of five subjects in our experiment choose to invest the 150 tokens in the binary choice problem and this is broadly similar across treatments. However, the table also reveals a strong gender effect. On average, men are much more inclined to choose the lottery than women are. The difference is highly significant (Fisher's exact test: p-value < 0.001).

We make a similar observation with respect to stage-2 investments (Table 4). On average, men allocate substantially more of their 300 tokens to the lottery than women do (p-value < 0.001). This confirms Hypothesis 1 (Gender) and leads to our

Result 1. Women invest less in the continuous token allocation task than men and are less likely to take the lottery in the binary choice task than men.

Moreover, Table 4 suggests a treatment effect. It appears that at least men, and to a lesser extent women, invest more in the RESP treatment than in RESUME or INHERIT. This observation already hints at our main research questions, and we will examine such differences in investment behavior more carefully further below.

Overall, the average allocation of tokens between the two options seems relatively moderate. Extreme investments consistent with the notion of risk (and loss) neutrality do not appear to be the norm. This is also largely confirmed by Fig. 5 which shows the distribution of investments across treatments and conditional on the stage-1 signal. Specifically, in RESUME and INHERIT we observe pronounced shifts in the distributions towards smaller investments when the signal is negative rather than positive. Very large

[15] Subjects were always paid exactly according to their accumulated earnings from the experiment except that the final amount was rounded up to the next 10 cents (e.g. €7.83 became €7.90). The rounding procedure was not announced in advance.

Table 2
Overview of experimental treatments.

	RESP	RESUME	INHERIT
Stage 1: Binary investment choice	Determination of PSP. Subject decides herself.	Predecessor's PSP reused for stage 1 and 2. Subject only observes predecessor's choice.	Predecessor's PSP reused for stage 1 and 2. Subject only observes predecessor's choice.
Lottery outcome	"Success" or "failure" feedback independently of stage-1 choice in all three treatments.		
Subject's payoff	Affected by outcome.	Not affected by outcome.	Affected by outcome.
Stage 2: Continuous investment	Current subject decides herself in all three treatments.		
Stage 3: Binary investment choice	n/a	Determination of a new PSP. Current subject decides.	Determination of a new PSP. Current subject decides.
Feedback	Stage-2 outcome and PSP are revealed.	Stage-2 and stage-3 outcomes, predecessor's PSP and own PSP are revealed.	Stage-2 and stage-3 outcomes, predecessor's PSP and own PSP are revealed.

Notes: PSP = Personal Success Probability.

Table 3
Share of subjects choosing the lottery in the binary-choice task.

	Overall	Females	Males
RESP	78.2%	59.1%	90.9%
RESUME	86.8%	68.8%	100.0%
INHERIT	82.2%	72.2%	88.9%
Total	81.9%	66.1%	92.7%

Table 4
Mean investments in the stage-2 token allocation task*.

	Overall	Females	Males
RESP	198.3	139.1	237.7
	(90.7)	(69.0)	(82.2)
RESUME	157.8	135.6	173.9
	(88.6)	(70.0)	(98.5)
INHERIT	160.4	125.4	183.7
	(85.7)	(62.0)	(92.3)
Total	174.8	133.7	202.8
	(90.0)	(66.2)	(93.6)

* Standard deviations in parentheses.

investments are rare in this case. In contrast, RESP subjects choose very large investments relatively frequently even in the negative-signal case, and the difference between the positive-signal and negative-signal distributions in RESP is more ambiguous than in the other treatments.

In RESUME and INHERIT it turns out to be important, as suspected, that subjects' own preferences regarding the binary investment choice are controlled for: Those who take the lottery in stage 3 invest on average 171.3 tokens in stage 2 while those who prefer the safe option invest only 94.2 tokens. This difference is statistically significant (p-value = 0.003). Thus, a RESP subject's stage-2 investment after that subject has chosen the lottery in stage 1 should only be compared to RESUME or INHERIT subjects' investments when those subjects have themselves likewise displayed a preference for the lottery in the binary choice task.[16] We will therefore adhere to this procedure when we make treatment comparisons further below.

4.2. Responses to positive versus negative signals in RESP

As discussed in Section 3.2, both expected utility theory and prospect theory predict that RESP subjects who choose the safe option in stage 1 will respond strongly to the signal, and in particular that the signal "failure" should cause these subjects to invest next to nothing in stage 2 (Hypotheses 2b and 3b). The data does not corroborate these predictions. Following the outcome "success" these more cautious participants invest on average 121.4 tokens. This is in principle compatible with predictions from both theories. However, following the outcome "failure" they still invest 111 tokens, far more than theoretically expected and not indicating any statistically significant difference between the two signals (p-value = 0.929). It is possible that this reflects insufficient Bayesian

[16] Note also that in RESP, too, there is a statistically significant difference in investments between lottery choosers and safe-option choosers (p-value < 0.001).

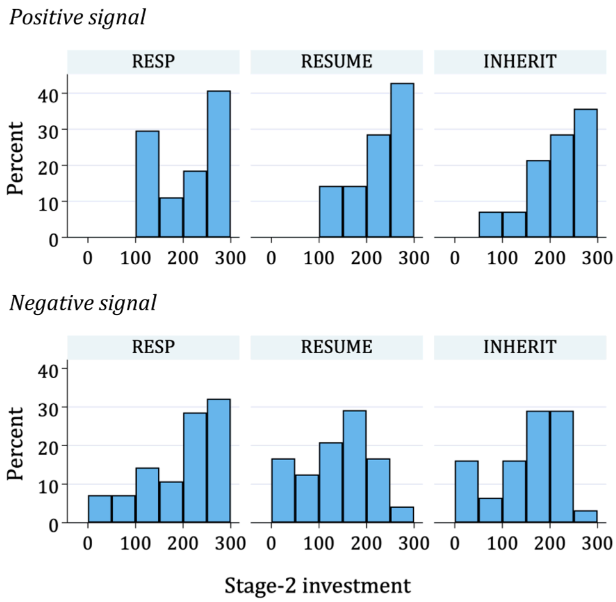

Fig. 5. Histograms of investments in the stage-2 token allocation task.

updating, a point which we will discuss in a little more detail in Section 4.3 below.

Next, we turn our attention to RESP subjects who select the stage-1 lottery. Do they react to a negative signal by going for lower investments than in the case of a positive signal, or do they display the contrarian response pattern as in Staw (1976)? As it turns out, they do neither. While they invest, overall, about a hundred tokens more than the subjects who prefer the safe option in stage 1, they too do not seem to respond very strongly to the signal. This is illustrated in Fig. 6. The average investment in this group following the outcome "success" is 237.5. This is in accordance with Hypothesis 3(a) and in particular with Hypothesis 2(a) which suggests that average investments in this case should exceed 150 tokens (p-value < 0.001 according to a two-sided *one*-sample Fisher-Pitman test). The corresponding figure following the outcome "failure" is 206.5 tokens, a little lower than the investment subsequent to a positive outcome but nowhere near the "minimal levels" predicted in Hypothesis 2(b). In fact, we cannot reject the null hypothesis that investments are the same after a positive or negative signal (p-value $= 0.245$).

Result 2. (a) Following the event "success" RESP subjects who choose to invest their 150 tokens in the stage-1 lottery invest even more in the stage-2 lottery while subjects who choose the safe option in stage 1 invest more cautiously. These results are consistent with Hypothesis 2 (RESP EUT) and Hypothesis 3 (RESP PT). (b) However, in contradiction to Hypothesis 2 investments remain largely unchanged when the stage-1 outcome is "failure". For subjects who prefer the safe option this is also at odds with Hypothesis 3, and for subjects who prefer the risky option the data does not confirm Hypothesis 8 (responsibility-based contrarian signal response).

Thus, the data is to some extent compatible with both expected utility theory and prospect theory but contradicts them in other ways. Expected utility theory is very successful in predicting differences in investments between subjects who make different choices in stage 1 and in predicting some levels, while the prospect theory model is less clear on this. Problematic for both theories is the lack of signal-specific investment behavior of those who prefer the safe option in stage 1. On the other hand, prospect theory is compatible with the finding that subjects who prefer the risky option in stage 1 do not react systematically to the signal (Hypothesis 3(c)), while this is completely at odds with expected utility theory. Finally, we do not find the contrarian signal response that Staw observed.[17]

4.3. The RESUME treatment

As mentioned above, an alternative explanation for the weak signal sensitivity of investments in RESP could be that our subjects underestimate the strength of the signal. They may conjecture intuitively that a single random draw contains hardly any information on the success probability. In combination with expected utility theory, such insufficient probability updating could account for our main results thus far. In combination with prospect theory, one might expect more escalation from subjects who suffer a loss but only partially consider the bad signal although this depends on the parameter constellations.

Our RESUME treatment acts as a control here because it removes any reasonable motive for escalation. The decision maker is

[17] According to prospect theory it could be that our subjects' preferences are quite homogeneous and characterized by parameter values that happen to be relatively close to the point of intersection—as illustrated in Figure 3—where investments subsequent to a gain versus a loss are the same, or it could be that some of our subjects escalate in an extreme way in response to an adverse outcome while others do not do so at all such that the choices of these two groups roughly balance out each other.

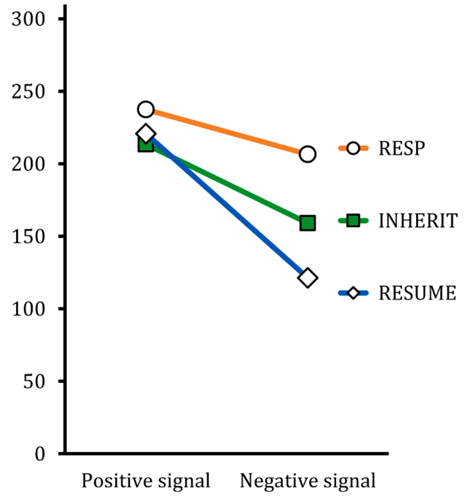

Fig. 6. Mean number of tokens invested in stage 2*
*Data based on subjects who choose the lottery in the binary choice task.

neither responsible for the stage-1 decision of picking the lottery, nor does she suffer a loss from a bad outcome. Yet, as seen in Fig. 6, even RESUME subjects invest sizeable shares of their 300 tokens when the lottery outcome suggests a low personal success probability, which is in contrast to the theoretical results.

Again, we might ask whether insufficient Bayesian updating could be behind the limited response to the signal. While this may be the case, note that such an approach does not work in the prospect theory model when we include the RESUME treatment in the analysis. Specifically, the model cannot explain mid-range investments after negative feedback in both treatments simultaneously. In the RESUME setting decision makers are predicted to invest positive amounts only if they are sufficiently optimistic about their chances of winning the second lottery after a bad signal from the first. However, in the RESP setting this level of optimism would then imply an escalation to the maximum investment.[18] Thus, depending on how subjects update the probability of success we should observe either extremely high stage-2 expenditures in the RESP treatment or zero stage-2 expenditures in the RESUME treatment. This is clearly not compatible with the moderate investment levels we observe.

Expected utility theory, on the other hand, would allow the possibility of moderate stage-2 behavior in both treatments if there is insufficient probability updating, but it does not predict treatment differences. Yet, we do find such differences. Although the average investment level after a bad stage-1 outcome is high even in RESUME, it is still systematically lower than in RESP (121.3 vs. 206.5; p-value = 0.003). Thus, the comparison between the RESUME treatment and the RESP treatment indicates the presence of escalation of commitment in the latter. In contrast, for the case of positive feedback, we cannot reject the null hypothesis that investments in the two treatments are the same.[19]

A further result is that subjects, unlike in the RESP treatment, do react systematically to the signal (positive: 220.7 tokens vs. negative: 121.3 tokens; p-value = 0.001). To examine more closely whether signal sensitivity diverges between treatments we follow a difference-in-difference approach based on a regression of the stage-2 investments on treatment dummies, a signal dummy, and their interactions. As before, our analysis focuses on subjects who select the lottery in the binary choice task, and we also control for gender. The results are reported in Table 5.

The regression analysis largely confirms the previous results and the visual impression from Fig. 6. First, subjects in the RESUME treatment, despite having received negative feedback (the reference category in Table 5), invest considerable amounts in the lottery option. Second, however, they do not reach the level observed in the RESP treatment. Third, switching to positive feedback leads to a further substantial and statistically significant increase in RESUME. Fourth, the large negative coefficient for the interaction between signal and RESP treatment indicates that changing the feedback makes much less of a difference in RESP. In the regression the difference in signal sensitivity between the two treatments is marginally significant.[20]

Result 3. Average stage-2 investments are lower in RESUME than in RESP. This is predominantly driven by a substantial reduction in investments following a negative stage-1 outcome (escalation of commitment). These findings are in line with Hypothesis 5

[18] Following a bad signal, the condition for positive investments when $D_2 = 0$ and for maximum investments (300 tokens) when $D_2 = -75$ is $p_L \geq \lambda_i/(2+\lambda_i)$ in both cases (where λ_i is the loss aversion parameter and p_L is the perceived or objective posterior probability of success; see Appendix B2 and B3).

[19] In fact, this is true for all pairwise comparisons among all three treatments. A Kruskal-Wallis test for the case of positive feedback delivers a p-value of 0.642.

[20] Testing for treatment differences between RESUME and RESP when the signal is positive yields non-significant results (effect size: 17.64; standard error: 26.77; p-value: 0.512).

Table 5
Regression of stage-2 investments on signal and on RESP vs. RESUME.

Positive signal	94.45***
	(27.11)
RESP	81.37***
	(23.85)
Pos. signal × RESP	–63.73*
	(35.86)
Male	59.81***
	(19.00)
Constant	83.54***
	(21.32)
Observations	76
R-squared	0.3338

†Standard errors in parentheses. Significance at the 10%, 5%, 1% level denoted by *, **, and ***, respectively.

(RESUME PT) and Hypothesis 7 (Escalation due to self-justification), whereas Hypothesis 4 (RESUME EUT) is rejected. Furthermore, RESUME investments are systematically affected by the signal, which is consistent with theoretical expectations, but they remain too high from a theoretical perspective when the stage-1 outcome is “failure”.

Just like Staw and in line with the idea of self-justification (Hypothesis 7) we find that when the feedback from the first lottery is *positive* it makes essentially no difference to investments whether the individual is responsible for the stage-1 decision or not. Evidently, our results differ from Staw’s in that *negative* feedback reduces investments in our RESUME treatment whereas this is not the case in Staw’s low personal responsibility treatment. Furthermore, in our RESP treatment negative feedback appears to have no effect while in Staw’s high personal responsibility treatment investments increase. A straightforward explanation for these patterns is that the contrarian response to feedback in Staw’s experiment is an artifact of his design. It appears that, with the message from the intermediate report at the beginning of stage 2 being rather ambiguous, the average subject in his experiment does not interpret bad news as an indication that their chosen path is really “failing”. Our results suggest that when there is a signal that contains more tangible information about the future profitability of the chosen path, decision makers do not escalate their commitment to a failing course of action to such an extent that they become more invested in it than if they had obtained good news. Nevertheless, the directional differences in subjects’ investments following negative feedback in RESUME and RESP are qualitatively the same as in Staw (1976). The open question is whether this is due to being (not) responsible for the stage-1 decision or due to having (not) incurred a loss.

4.4. The inherit treatment

Our final treatment reintroduces the financial consequences from “success” or “failure” without endowing the stage-2 decision maker with the responsibility for the path taken in stage 1. If responsibility is irrelevant there should be no difference between the INHERIT and the RESP treatment, which is what both EUT and prospect theory predict.

Our results are mixed. Consider Fig. 6 and Table 6 in which we extend our earlier regression analysis on treatments and signals. We find that investments following negative feedback in INHERIT are higher than in the RESUME treatment but lower than in the RESP treatment. These differences are similar in size and both appear to be systematic (although the RESUME variable in Table 6 is only weakly significant).[21] Thus, we still detect some escalation of commitment but it is more limited than in the RESP treatment.

The interaction effects from the difference-in-difference analysis also suggest that we can think of INHERIT behavior as being situated somewhere in between the other two treatments: Quantitatively, the impact of the signal on investments is somewhat stronger than in RESP but somewhat weaker than in RESUME. However, neither of these differences is statistically significant.

Result 4. Not being responsible for the stage-1 decision affects the stage-2 investments when the stage-1 outcome is negative and leads to lower investments in INHERIT than in RESP. Furthermore, INHERIT subjects invest systematically less after a negative signal than after a positive one. These results are in line with Hypothesis 7 (Escalation due to self-justification) while Hypothesis 6 (INHERIT EUT-PT) is rejected.

Overall, our main treatment comparisons are best explained by a combination of prospect theory’s reflection effect and a responsibility effect due to self-justification. It appears that when subjects experience a monetary loss in stage 1, at least some of them want to grab the opportunity to make up for it in stage 2 but the motivation to act in this way is stronger when some personal responsibility for that loss can be attributed to them.

4.5. Gender effects

As reported above, women display more risk aversion (or loss aversion) than men, as hypothesized. In addition, we observe further differences between male and female participants in our data that we did not expect to find. Specifically, for women, unlike for men,

[21] We obtain virtually the same statistical results when we use non-parametric tests.

Table 6
Regression of stage-2 investments on signal and treatment.

Positive signal	49.94**
	(25.13)
RESP	42.78**
	(21.30)
RESUME	–38.20*
	(22.39)
Pos. signal × RESP	–19.25
	(33.59)
Pos. signal × RESUME	44.01
	(35.91)
Male	65.89***
	(14.66)
Constant	117.90***
	(17.47)
Observations	113
R-squared	0.3412

†Standard errors in parentheses. Significance at the 10%, 5%, 1% level denoted by *, **, and ***, respectively.

there are no clear treatment effects and there are no statistically significant differences in investments following "success" versus "failure" in any treatment, not even in RESUME.

It is important to point out that our experiment was not designed to study treatment-specific gender effects and that we had not formulated any relevant hypotheses in this respect before conducting the experiment. Furthermore, there is a problem of statistical power as we have fewer female than male participants to begin with and fewer females choose the lottery in the binary-choice problem (see Result 1). Our data is therefore not suitable to draw any firm conclusions about gender-specific behavior in the context of the setting we study. However, the differences in the behavioral patterns are sufficient to persuade us that future research on gender effects in escalation of commitment may be warranted. To our knowledge not much research of this kind currently exists. One study we are aware of (Lam and Ozorio, 2013) considers gambling decisions of males and females after a win or a loss in a casino-type task. The authors report mixed results but at least for one subgroup they found some increased tendency for women to "escalate". However, the context was different. For example, wins and losses were determined as i.i.d. draws and thus contained no information on the probability of success.

5. Conclusions

Our paper contributes to the thus far relatively small experimental literature on the sunk cost fallacy in the presence of monetary incentives. We examine a setting that has been very influential in psychology and organizational behavior research and has become a workhorse for investigating the phenomenon of escalation of commitment to a failing course of action. We find that subjects respond differently to negative feedback about a previous investment when this feedback is linked to a financial loss or associated with personal responsibility for the earlier investment. This is compatible with two leading explanations for escalation of commitment, prospect theory and the notion of self-justification due to cognitive dissonance. However, unlike Staw (1976) we do not find increased expenditures relative to the reaction to positive feedback.

An important difference between our setting and Staw's is that our feedback at the beginning of the second stage contains a meaningful signal about the profitability of the activity that decision makers have become engaged in. The presence of such a signal is vital though, if one is interested in making inferences about the rationality of the actions that subjects take. A critic might point out that our signal is too strong to be overshadowed by escalation. After all, if it is crystal-clear that a chosen path is not likely to lead to a good outcome, additional commitment can plausibly only be based on a severe misinterpretation of the situation since there is then little scope for "sugarcoating" the earlier decision. However, the limited response of our subjects to the signal, particularly in the RESUME treatment, suggests that there is in fact sufficient uncertainty to allow a rationalization of investments in the case of negative feedback even when decision makers are not risk-prone.

Although we are unable to replicate all aspects of Staw's results, our findings do suggest a reference-dependent shift in risk attitudes causing decision makers to respond in a non-intuitive way (from a rational-choice perspective) to the signal. They also suggest that follow-up investment decisions can be affected by whether the decision maker has been responsible for a relevant preceding choice. While the former can be incorporated into economic theory via prospect theory and its potential successors (see, for example, Kőszegi and Rabin, 2007), it is less clear how to model economic behavior that is contingent on personal responsibility. We think that this poses an interesting challenge.

It is, however, possible that our results under- or overestimate the effects of responsibility in naturally-occurring applications in which we might suspect escalation of commitment to occur. The reason is that there may be many such applications in which the decision maker, unlike in our setting, acts on the behalf of others and may have to answer for the consequences of his or her choices. On the one hand, it could be that this makes the individual more inclined to appear consistent and stick to a chosen course of action. On the other hand, it is also conceivable that the individual has a greater incentive to learn to avoid such biases when these ultimately lead to harmful outcomes. Exploring these issues is an interesting avenue for future research in our view.

Acknowledgements

We thank two anonymous reviewers for very thoughtful comments. Furthermore, we are very grateful for excellent research assistance by Felix Holub, Katharina Momsen, Marcel Preuß and Raphael Epperson. Thanks also for valuable comments from seminar participants at the ESA European Meeting in Heidelberg and at the Annual Meeting of the Standing Field Committee of the German Economic Association 2017 in Göttingen.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.euroecorev.2021.103811.

References

Berg, J.E., Dickhaut, J., Kanodia, C., 2009. The role of information asymmetry in escalation phenomena: empirical evidence. J. Econ. Behav. Organ. 69, 135–147.
Brockner, J., 1992. The escalation of commitment to a failing course of action: Toward theoretical progress. Acad. Manage. Rev. 17 (1), 39–61. https://doi.org/10.2307/258647.
Buchheit, S., Feltovich, N., 2011. Experimental evidence of a sunk-cost paradox: a study of pricing behavior in Bertrand–Edgeworth duopoly. Int. Econ. Rev. (Philadelphia) 52 (2), 317–347.
Cachon, G.P., Camerer, C.F., 1996. Loss-avoidance and forward induction in experimental coordination games. Q. J. Econ. 111 (1), 165–194.
Camerer, C.F., Weber, R.A., 1999. The econometrics and behavioral economics of escalation of commitment: a re-examination of Staw and Hoang's NBA data. J. Econ. Behav. Organ 39 (1), 59–82.
Charness, G., Gneezy, U., 2012. Strong evidence for gender differences in risk taking. J. Econ. Behav. Organ 83 (1), 50–58.
Croson, R., 2005. The method of experimental economics. Int. Negot. 10 (1), 131–148.
Feltovich, N., Iwasaki, A., Oda, S.H., 2012. Payoff levels, loss avoidance, and equilibrium selection in games with multiple equilibria: an experimental study. Econ. Inq. 50 (4), 932–952.
Festinger, L., 1957. A Theory of Cognitive Dissonance. Stanford University Press.
Friedman, D., Pommerenke, K., Lukose, R., Milam, G., Huberman, B.A., 2007. Searching for the sunk cost fallacy. Exp. Econ. 10, 79–104.
Gneezy, U., Potters, J., 1997. An experiment on risk taking and evaluation periods. Q. J. Econ. 112, 631–645.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Haita-Falah, C., 2017. Sunk-cost fallacy and cognitive ability in individual decision-making. J. Econ. Psychol. 58, 44–59.
Hertwig, R., Ortmann, A., 2001. Experimental practices in economics: a methodological challenge for psychologists? Behav. Brain Sci. 24 (3), 383–403.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263–291.
Kanodia, C., Bushman, R., Dickhaut, J., 1989. Escalation errors and the sunk cost effect: an explanation based on reputation and information asymmetries. J. Account. Res. 27 (1), 59–77.
Köbberling, V., Wakker, P.P., 2005. An index of loss aversion. J. Econ. Theory 122 (1), 119–131.
Kőszegi, B., Rabin, M., 2007. Reference-dependent risk attitudes. Am. Econ. Rev. 97 (4), 1047–1073.
Lam, D., Ozorio, B., 2013. The effect of prior outcomes on gender risk-taking differences. J. Risk Res. 16 (7), 791–802.
Leeson, N., 1996. Rogue trader: How I brought Down Barings Bank and Shook the Financial World. Little, Brown & Company, New York.
Phillips, O.R., Battalio, R.C., Kogut, C.A., 1991. Sunk and opportunity costs in valuation and bidding. South Econ. J. 58 (1), 112–128.
Rydval, O., Ortmann, A., 2005. Loss avoidance as selection principle: evidence from simple stag-hunt games. Econ. Lett. 88, 101–107.
Sleesman, D.J., Conlon, D.E., McNamara, G., Miles, J.E., 2012. Cleaning up the big muddy: a meta-analytic review of the determinants of escalation of commitment. Acad. Manage. J. 55 (3), 541–562.
Sleesman, D.J., Lennard, A.C., McNamara, G., Conlon, D.E., 2018. Putting escalation of commitment in context: a multilevel review and analysis. Acad. Manage. Ann. 12 (1), 178–207.
Schulz-Hardt, S., Thurow-Kröning, B., Frey, D., 2009. Preference-based escalation: a new interpretation for the responsibility effect in escalating commitment and entrapment. Organ. Behav. Hum. Decis. Process. 108 (2), 175–186.
Staw, B.M., 1976. Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action. Organ. Behav. Hum. Perform. 16 (1), 27–44.
Staw, B.M., and J. Ross (1987). Behavior in escalation situations: antecedents, prototypes, and solutions. In: B.M. Staw and L.L. Cummings (Eds.), *Research in Organizational Behavior* 9, 39–78. Greenwich, Connecticut: JAI Press.
Staw, B.M., Ross, J., 1989. Understanding behavior in escalation situations. Science 246 (4927), 216–220.
Thaler, R., 1980. Toward a positive theory of consumer choice. J. Econ. Behav. Organ. 1 (1), 39–60.
Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. J. Risk Uncertain. 5 (4), 297–323.
Weber, M., Zuchel, H., 2005. How do prior outcomes affect risk attitude? Comparing escalation of commitment and the house-money effect. Decision Analysis 2 (1), 30–43.
Whyte, G., 1986. Escalating commitment to a course of action: a reinterpretation. Acad. Manage. Rev. 11 (2), 311–321.