Computers in Human Behavior 133 (2022) 107270

Contents lists available at ScienceDirect

Computers in Human Behavior

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

# When humans and computers induce social stress through negative feedback: Effects on performance and subjective state

S. Thuillard [a,*], M. Adams [a], G. Jelmini [a], S. Schmutz [c], A. Sonderegger [b], J. Sauer [a]

[a] Université de Fribourg, Rue P.- A. de Faucigny, 1700 Fribourg, Switzerland
[b] Bern University of Applied Sciences, Business School, Institute for New Work, Brückenstrasse 73, 3005 Bern, Switzerland
[c] Inselspital Bern, Universitätsspital Bern, Freiburgstrasse, 3010 Bern, Switzerland

ARTICLE INFO

*Keywords:*
6 max): Social stress
Negative feedback
Performance
Computer feedback
Automation
Interpersonal fairness

ABSTRACT

People increasingly work with autonomous systems, which progressively take over functions previously performed exclusively by humans. This may lead to situations in which automated agents give negative performance feedback, which represents an important work-related social stressor. Little is known about how negative feedback provided by computers (as opposed to humans) affects human performance and subjective state. A first experiment (N = 60) focused on the influence of human feedback on performance. After participants had performed a cognitive task, they received a manipulated performance feedback (either positive or negative) from a human (comparing to a control with no feedback) and subsequent performance on several cognitive tasks and the participants' subjective state was measured. The results showed that while negative feedback had a negative influence on several subjective state measures, performance remained unimpaired. In a second experiment (N = 89), participants received manipulated negative feedback by a human or by a computer (or no feedback at all) after having completed an ability test. Subsequent performance was measured on attention tasks and creativity tasks and participants' subjective state was assessed. Although participants felt stressed by both negative computer and human feedback, subsequent performance was again not impaired. However, computer feedback was rated as being less fair than human feedback. Overall, our findings show that there are costs of protecting one's performance against negative feedback and they call for caution regarding the use of negative feedback by both human and automated agents in work settings.

## 1. Introduction

Humans at work may be exposed to different social stressors, such as bullying, ostracism, harassment, or negative performance feedback. These social stressors refer to different types of interactions between employees (at different hierarchical levels or not), which can affect a person's social esteem and self-esteem by initiating cognitive evaluative processes (Semmer et al., 2007). This may have serious implications for employees at the psychological, physical and behavioral levels (Semmer et al., 2019).

Of these social stressors, negative performance feedback (i.e. informing someone of her or his inadequate performance), is particularly prevalent in work settings (Cleveland et al., 1989; Sauer et al., 2019). In addition to the formal and planned procedure of providing performance feedback as part of an organizational appraisal process, the prevalence of the stressor can be increased when supervisors give spontaneous comments on the performance of an employee that (without the supervisor being aware) may contain subtly offending cues (Krings et al., 2015). While the effects of human feedback have been studied in some depth (see for example Cawley et al., 1998; Kuvaas, 2006; Stanton, 2000), the literature remains rather inconclusive regarding its influence on human performance. In this regard, previous research has shown that negative feedback may improve, impair, or not affect subsequent performance. It has been argued that expecting simple causal relationships between negative feedback and performance would be too simplistic (Ilgen et al., 1979). Instead, additional factors such as characteristics of the feedback, of its source and of its recipient may play a role. Furthermore, experimental research on negative feedback may potentially overlook or confound the distinction between feedback source and feedback medium (see section 1.2).

While performance feedback and social stress were until now exclusively addressed in contexts of human-human collaboration (or

* Corresponding author.
*E-mail address:* simon.thuillard@unifr.ch (S. Thuillard).

https://doi.org/10.1016/j.chb.2022.107270
Received 9 October 2020; Received in revised form 27 December 2021; Accepted 12 March 2022
Available online 16 March 2022

leadership), these topics received recently increased attention in the context of human-automation interaction (Sauer et al., 2019). Automation and technological advancements allowed computers and machines to take over managerial tasks (Wesche & Sonderegger, 2019), including providing negative performance feedback (Sauer et al., 2019). With these developments arises the question of whether computer feedback[1] affects its recipients differently from human feedback. Automated negative feedback to human employees has been used in organizations for some years now. One well-known application is in the transport company Uber, in which drivers are managed by an automated system that allocates tasks, plans shifts, and gives performance feedback (Wesche & Sonderegger, 2019). The literature review on the effects of computer feedback on performance (again focusing on experimental studies using objective measures of performance), highlighted a similar variety of results as with human feedback. Again, these differences might be related to different feedback characteristics, though we think another explanation is possible.

In this matter, we raise the question regarding the transferability of older research findings of computer feedback to more modern forms of computer feedback. The literature we reviewed on negative computer feedback spans several decades, from 1985 to 2016. Perceptions of and attitudes towards technology are very likely to have changed over time, as technology itself evolved considerably (Alder & Ambrose, 2005). Such a change in the perception of computers over time has been observed (Gardner et al., 1989; Immonen & Sintonen, 2015). In the field of algorithmic reliance, recent research showed that participants followed advice more frequently when they thought it came from an algorithm rather than a human (Logg et al., 2019). This result was different from research conducted in previous years, which often found aversion towards algorithmic advice (e.g., Dietvorst et al., 2015; Dzindolet et al., 2002; Promberger & Baron, 2006). Overall, these examples show that perceptions of and attitudes towards technology can shift with time and technological progress. This shift may take place as well in relation to computer feedback, suggesting that previous and current forms of computer feedback are different to such an extent that they have different effects. Therefore, we argue that research needs to be conducted on modern forms of automated feedback (for example based on algorithms or deep learning) to obtain more ecologically valid results with regard to the impact of current technologies. This was done in the second study of this article.

The main goal of this article is to understand how negative performance feedback as a form of social stress affects recipients' behavior (subsequent performance) and subjective reactions, when induced by either a human or a computer. This was done in two parts. Study 1 evaluated the effect of negative human performance feedback on subsequent performance on a wide range of cognitive tasks and on subjective state, compared to no feedback or positive feedback, while also putting the effectiveness of our experimental manipulation to the test. In Study 2, we investigated the effect of modern forms of automated feedback provided by a computer agent, compared to negative human feedback and no feedback, still using a wide range of cognitive tasks and assessing additional subjective constructs. Study 2 should contribute to the literature as it is the first one to investigate a *modern* form of automated feedback, and it does so by focusing on computers as the source of feedback and not the medium (see section 1.2 for an explanation of this distinction).

The studies depicted in the current paper were exploratory in design and intent. The literature is inconclusive regarding the link between negative feedback provided either by humans or technology/computers on performance. Thus, we considered our work to be exploratory in nature, allowing us to investigate a broad range of variables, possibly finding some leads for future research. Additionally, there is to our knowledge no previous experiment investigating a modern form of computer feedback which we could use to guide our study. Implications of this exploratory nature are threefold. First, no a priori power analyses were conducted, though Study 2 aimed at groups large enough ($N = 30$) to have a normal distribution. Second, variables in Study 1 and Study 2 differed to some degree. On the one hand, this allows covering a large horizon of variables, congruently with our exploratory design. On the other hand, some choices had to be made due to time constraints for the length of experiments as well as the development process of the project. On that last point, in the time gap between Studies 1 and 2, more literature and concepts came to our attention which we decided to include in Study 2. Details on which variables exactly were added for Study 2 are presented in section 1.4. Third, some variables were added to the experiments because they were of potential interest, even though there was sometimes a lack of literature about these variables with regards to social stress. The theoretical background below details the literature review that was conducted on the effects of human and computer negative feedback on performance. We then present the main dependent variables used in the two studies, as well as relevant theoretical models.

### 1.1. *Performance feedback as a stressor*

Although commonly used as a technique in human resource management, performance feedback is quite often a source of dissatisfaction and stress for both employees and supervisors (Fletcher, 1997; Murphy & Cleveland, 1995). During an appraisal interview, negative performance feedback may result in undesirable consequences at several levels (Holbrook, 2002). This is particularly true when the feedback is too general, inconsiderate, contains threats and attributes poor performance to internal factors. Such features are typical for destructive negative feedback, which can have strong effects on the recipient (Baron, 1988). In the current article, both "standard" negative feedback and destructive negative feedback were examined, in Study 1 and Study 2 respectively.

Sauer et al. (2019) recently proposed three mechanisms to explain how social stress (including negative performance feedback) can affect performance. (1) 'Blank out'-mechanism: despite being the target of social stress, the employee is able to protect his or her performance from being impaired. If this mechanism takes effect, nil effects on performance will be found. This mechanism may take place when safety-critical tasks are carried out, representing a context in which decreased performance needs to be avoided because it might have serious safety-related consequences. (2) 'Rumination'-mechanism: due to negative feedback threatening his or her self-esteem, the employee ruminates about it. It is expected that such thoughts distract cognitive resources from the main task, leading to impaired performance. This mechanism is similar to what was proposed in the Feedback Intervention Theory (Kluger & DeNisi, 1996). (3) 'Increased motivation'-mechanism: the employee responds to negative feedback by trying to demonstrate that he or she can do better than that, leading to an improvement of subsequent performance. This mechanism may take effect when the employee performed the task at reduced levels of motivation and effort expenditure. In this case, self-esteem of the employee is not threatened by the social stressor. Following negative feedback, the employee may then decide to increase effort expenditure resulting in improved performance. The three mechanisms are part of the theoretical framework of this article, helping to improve our understanding of the effects of negative feedback on performance.

#### 1.1.1. *Human feedback*

Research on how feedback affects subsequent performance has a long tradition, going back to the beginning of the last century (Kluger & DeNisi, 1996). Being influenced, for example, by Ammons' review (1956), there was for a long time a consensus in the research literature that any type of feedback (even if negative) would improve

[1] Please note that in this article, the terms computer feedback or automated feedback are used as general terms referring to any type of technology that can communicate with humans and that is used in a work environment.

performance, typically by increasing learning and motivation. When Kluger and DeNisi (1996) reviewed the literature, they identified biases and methodological problems with previous research, largely demonstrating that the idea of general feedback always improving subsequent performance was false. Instead, they found a large variety of effects of feedback in their meta-analysis. While feedback interventions often improved performance, it sometimes had no effect or even impaired performance in about one third of cases. A considerable number of moderating variables of the effect of feedback on performance were identified in the meta-analysis. For example, according to the moderator analyses by Kluger and DeNisi (1996), feedback tended to impair performance if it praises, discourages, threatens self-esteem or is given verbally. Conversely, performance tended to improve if feedback contained the correct solution, informed about the change of performance since last feedback, or was delivered by a computer.

We continued the review of the literature on human negative feedback posterior to Kluger and DeNisi's meta-analysis, focusing on experimental studies with objectively measured performance. We chose these selection criteria to focus on the most relevant literature for this present article, since it is concerned with experimental studies measuring performance objectively. The goal was to examine whether such a variability of results would be found again in this specific part of the literature. The review revealed that further experiments have confirmed Kluger and DeNisi's claim that negative feedback can indeed impair subsequent performance (Alder, 2007; Alder & Ambrose, 2005; Nease et al., 1999; Raver et al., 2012). This raises some concerns that appraisal interviews would fail their purpose if negative performance feedback would actually lead to performance decreases rather than increases (considering that performance improvement represents one of the main goals of the appraisal process; Holbrook, 2002). However, this was not the case in all studies. For example, Alder (2007) found performance in a clerical task to be improved following constructive negative feedback from a supervisor. Finally, recent work also suggested the occurrence of nil effects on objective performance measures in a highly complex task environment (Peifer et al., 2020). This may be due to the complexity of the task used in this experiment, as more complex tasks have been found to reduce the effect of feedback interventions (Kluger & DeNisi, 1996). Overall, although most studies found subsequent performance to be impaired, performance improvements and nil effects were also found. These results support the findings of Kluger and DeNisi (1996), showing that negative human feedback can affect performance in different directions.

#### 1.1.2. *Computer feedback*

An even larger variety of effects is found in the literature on computer negative feedback on performance. Kluger and DeNisi's meta-analysis (1996) showed that computer feedback in general improved performance. However, focusing again on experimental studies measuring performance objectively, negative computer feedback was found to lead to either improved performance (Alder, 2007; Earley, 1988; Fyfe & Rittle-Johnson, 2016; Nebeker & Tatum, 1993; Van Dijk & Kluger, 2011), decreased performance (Alder, 2007; Resnik & Lammers, 1985; Van Dijk & Kluger, 2011), or had no effect at all (Kluger & Adler, 1993; Sauer et al., 2020). Overall, our literature review shows that negative feedback provided by human and computer alike can affect performance in several ways.

This variety of results has already been discussed in the literature long ago by Ilgen, Fisher and Taylor (1979). They pointed out that assuming simple causal relationships between feedback and performance would be oversimplifying. Instead, characteristics of feedback, of its source and of its recipient should influence feedback effects on behavior and subsequent performance. A recent review by Lechermeier and Fassnacht (2018), focusing on feedback source, timing and valence, reiterated this point. They also stated that the main effects on performance are inconsistent, and that they can vary considerably depending on the source, message, task or individual characteristics. This might explain why so many different results patterns can be found when comparing the effects of human and computer feedback on performance. Relevant characteristics for the present studies are discussed below.

### 1.2. *Feedback source and task type*

To understand the respective effects of negative human and computer feedback, it is crucial to make a distinction between feedback source and medium (Alder & Ambrose, 2005). Source refers to the agent generating the feedback (i.e. human or computer) while medium refers to by whom the feedback is given (i.e. human or computer). This distinction may affect the extent to which subsequent performance is impaired. For example, face-to-face human feedback provides participants with an opportunity to justify themselves. This possibility of justifying oneself after receiving negative feedback can have positive effects on the recipient's reactions such as perceived interpersonal fairness, and possibly performance (Alder, 2007; Alder & Ambrose, 2005). The main implication of these results is that when examining the effect of feedback source alone, feedback medium needs to be the kept the same across experimental conditions. We noticed this was not the case in the literature we reviewed. Most studies focused on either human or computer feedback only (Nease et al., 1999; Nebeker & Tatum, 1993; Raver et al., 2012; Resnik & Lammers, 1985; Van Dijk & Kluger, 2011). Some of this work examined negative feedback but operationalized it as computer feedback without conceptualizing it as a specific source potentially having specific effects. One study looked at computer feedback and varied only the medium (Alder & Ambrose, 2005). Two studies compared human and computer negative feedback (Alder, 2007; Kluger & Adler, 1993). However, in both cases feedback source was manipulated as well as feedback medium. Human source was paired with human medium, and computer source was paired with computer medium. While these methods are clearly valid, the literature shows that no study so far has truly isolated the effect of feedback source. Study 2 is thus the first study to investigate feedback source alone, by keeping feedback medium constant across conditions.

Since this article focuses on objective performance as an important outcome measure, the role of task type needs to be addressed. Van Dijk and Kluger (2011) showed, for example, that the effect of negative computer feedback on performance was influenced by the type of task to be performed. The authors distinguished between two types of tasks based on the regulatory focus theory by Higgins (1997). This theory postulates that humans have two regulatory foci: the prevention focus that regulates goals of avoiding punishment, and the promotion focus that regulates goals of achieving rewards. Van Dijk and Kluger (2011) extended the notion of regulatory focus to task type, showing that some tasks would induce a promotion focus while others would induce a prevention focus. They thus made a distinction between prevention tasks (i.e. requiring error avoidance and caution such as in proofreading) and promotion tasks (i.e. requiring imagination and an open mind such as in product development). In prevention tasks, negative feedback improved subsequent performance, whereas positive feedback decreased it. In promotion tasks, subsequent performance decreased following negative feedback, and improved when positive feedback had been given. Additionally, in Kluger and DeNisi (1996) and Lechermeier and Fassnacht (2018), task type was found to moderate the effect of feedback on performance for human feedback as well. A major implication of this work is that research on the effects of feedback should use different types of tasks to measure performance-related effects.

### 1.3. Subjective effects

In order to obtain a more complete picture of the effects of negative performance feedback on recipients, it appears insufficient to measure performance alone. Subjective reactions to feedback must be investigated as well. Subjective indicators of strain were classified by Sauer et al. (2019) as a group of outcome variables, along with performance, with which effects of social stress can be measured. Additionally, the model of compensatory control mechanism (Hockey, 1997) predicts active human performance management with a view to protecting overall task performance. This is typically in the form of taking some compensatory action that sometimes involves adaptations at the cognitive-energetical level (e.g., increased effort expenditure, increased focus on primary task). This illustrates how performance may be protected from negative feedback, but at the cost of a subjective strain that may be detected by subjective indicators and not by objective tests.

The 'Stress as Offense to Self' theory (SOS; Semmer et al., 2019) focuses on the effects of social stress on a person's subjective state and well-being. It postulates that social stress acts mainly through threats to the self. The SOS approach identifies three mechanisms in which social stress, and negative feedback in particular, might impinge on the self. (1) Stress as thwarting important goals: one almost universal goal is to maintain self-esteem, which can be threatened by receiving negative performance feedback. (2) Stress through insufficiency: one may feel inadequate following negative feedback. (3) Stress as disrespect: independently of the content, the way negative feedback is given can be offensive or disrespectful, and thus stressful. Based on this model, we added a measure of state self-esteem in Study 2 in order to verify whether social stress in the form of negative feedback does indeed act on the recipient's self-esteem. The SOS approach, through its mechanisms on social stress affecting well-being, constitutes an additional reason to use subjective variables in the present studies.

In line with assumptions of the SOS paradigm, previous research has indicated subjective consequences of negative feedback. For example, at the personal level negative feedback can lower one's feelings of self-worth (Brown, 2010) and impair self-esteem (Krings et al., 2015; Moore & Klein, 2008). At the emotional level, negative feedback can induce negative affective states such as anger or tension (Baron, 1988; Cianci et al., 2010), and even stress reactions such as anxiety (Nummenmaa & Niemi, 2004). At the relationship level, the negative feedback giver is more likely to be blamed and be less trusted by the recipient (Raver et al., 2012), and perceived interpersonal fairness can be impaired (Alder, 2007). This last construct refers to the degree of perceived fairness in the personal relationship between the feedback giver and the recipient (Colquitt et al., 2015). It was deemed a key construct by Alder (2007) in understanding the effect of negative feedback source on performance, and as such was added in Study 2.

### 1.4. Present studies and hypotheses

The main goal of the two studies was to examine how negative feedback as a prominent social stressor affects subsequent task performance and subjective state of the feedback recipient. This question is addressed for human and computer feedback, with a modern form of computer feedback being used in the latter case. The present work used different types of tasks to investigate performance-related effects of negative feedback. In the first study, we investigated whether positive and negative feedback provided by a human would have different effects on performance, using a wide range of established tasks measuring different types of cognitive performance, with a control group (i.e. no feedback) serving as a baseline. The second study focused on negative feedback, making a comparison between the two sources (i.e. human versus computer) while controlling for medium, and again a control group that received no feedback. This study used a somewhat different set of tasks than the first to increase the total number of tasks being examined and to use both promotion and prevention tasks as defined by Van Dijk and Kluger (2011). Based on our experience of the first study, the number of subjective state measures was increased to include further relevant concepts. More precisely, we added interpersonal fairness, level of distraction and motivation to improve based on Alder (2007), and state self-esteem based on the SOS (Semmer et al., 2019). Additionally, to measure affect we used a shorter questionnaire in Study 2 than in Study 1 due to time constraints in the second experiment.

The methodological approach used in both studies was similar in that it used previous lab-based manipulations of social stress under highly controlled conditions. In both studies, we employed cognitive tests that are well established in personnel selection and other diagnostic settings. In each study a slightly different set of tasks was performed. In the first study, we measured processing speed, perceptual reasoning, backward counting and attention while in the second one, we investigated two types of creativity and attention again. This allowed us to examine the effects of negative performance feedback on a large set of outcome measures. Overall the two studies are complementary since Study 1 was the basis on which we tested our experimental manipulation and procedure while also investigating several performance tasks. Study 2 extended this work by adding computer feedback as well as more tasks and measures while controlling for the effect of medium. Additionally, Study 2 used the same attentional performance task as Study 1, aiming at replicating the result. Three main research questions were addressed in these two studies. (a) Does negative performance feedback lead to poorer performance on tasks that are completed following the feedback? (b) Are different types of tasks affected to a different extent? (c) Does feedback generated by humans and computers impair subsequent performance differently?

In Study 1 no hypothesis on the performance variables was put forward due to the inconclusive research findings with regard to performance. Based on the assumption that performance appraisal involving feedback was generally stressful, we hypothesized regarding subjective variables that:

- 1a) Receiving negative and positive performance feedback will result in higher state anxiety than when receiving no feedback, with negative feedback showing the highest strain levels.
- 1b) Receiving negative feedback will induce higher negative affect than no feedback, and receiving positive feedback will induce the lowest negative affect. We expected reverse effects for positive affect.

In Study 2, performance-related hypotheses were based on the regulatory focus theory by Higgins (1997), which was extended to task type by Van Dijk and Kluger (2011). The latter article showed that negative feedback impaired performance in promotion tasks, such as creativity tasks (H2a). In prevention tasks such as attention tasks, performance should be improved following negative feedback. However, performance on the same attention task in Study 1 was not impaired by negative feedback. We thus formulated H2b based on this result of Study 1 instead of Van Dijk and Kluger (2011). As a social stressor, and following the SOS theory (Semmer et al., 2019), negative feedback should have a negative impact on subjective variables such as affect (H2c) or state self-esteem (H2d). The variables in H2e were based on Alder (2007). As explained in section 1.2, Alder (2007) compared human and computer negative feedback without controlling for medium. The results from this reference were not sufficiently strong to formulate directed hypotheses. However, it still helped provide a general expectation about these variables in H2e.

The specific hypotheses for Study 2 were as follows:

- 2a) Based on results by Van Dijk and Kluger (2011), performance in creativity tasks will be lower in the human and computer feedback conditions than in the control group.
- 2b) Attentional performance will not be affected by negative feedback.
- 2c) Overall affect will be more negative in the human and computer feedback conditions than in the control group.
- 2d) State self-esteem will be lower in the human and computer feedback conditions than in the control group.
- 2e) We generally expected feedback source to affect interpersonal fairness, level of distraction and desire to improve. However, since this is the first study to truly investigate feedback source alone, we could not formulate directed hypotheses for these variables.

## 2. Study 1

### 2.1. Goal of the study

The goal of the first study was to examine whether negative performance feedback induced by a human as a prominent social stressor at work can be modelled in a lab-based context, and whether it would impair cognitive performance on several subsequent tasks as well as subjective state measures.

### 2.2. Participants and experimental design

Sixty students from the University of Fribourg took part in the study (23 females, age between 19 and 53 years, $M = 24.12$, $SD = 5.31$). They were recruited from all university departments, except for the departments of psychology, education and special education. This was because students from these departments generally have some good knowledge of psychological testing and might have been less responsive to the experimental manipulation. For the same reason, we excluded participants who had previously completed an intelligence test because they may have known their personal test score. Half of the students were German native speakers, the other half were French native speakers. The experimental materials were available in both languages. Participants received CHF 20.- as a financial compensation for their participation.

A one-way between-subjects design was implemented in the experiment. The independent variable ‘social stress’ was manipulated through inducing feedback at three levels: positive performance feedback, negative performance feedback and no performance feedback.

### 2.3. Dependent variables

*Manipulation check.* The following three items were used as a manipulation check to verify whether the experimental manipulation was successful: (1) “*How do you evaluate your own performance on the test?*” (*very poor – very good*); (2) “*How much stress are you feeling right now?*” (*very little – a great deal*); (3) “*How stressful did you find the performance feedback to be?*” (*not at all – very*). Each item used a 10-point Likert scale. Participants completed the items after having received performance feedback. These questions where formulated in order to see whether the participants actually believed in the feedback they received and to assess whether negative feedback was actually stressful.

*Cognitive performance.* Four standardized tests were used to measure different facets of cognitive performance. (a) *Backward counting:* The participants were asked to count down from number 1022 in steps of 13 over a period of 150s, following a procedure adapted from the Trier Social Stress Test (Kirschbaum et al., 1993). Each time participants made a mistake, they were asked to start again from the beginning. The number of mistakes was used as an indicator of performance. (b) *Attentional performance:* attention and concentration performance was measured in a sustained visual scanning task, the d2-test (Brickenkamp, 1962), in the form of accuracy (errors in %) and speed (number of characters processed). Participants completed the first 10 lines of the test. (c) *Digit symbol:* This test of the WAIS-R (Wechsler, 1981) measures perceptual speed and visual-motor coordination. In this test, the association of nine symbols with their corresponding number (1–9) was shown to the participant. Participants were then presented a list of 93 symbols, with each of them having to be marked with the corresponding number (within a total time limit of 90s). Performance was scored by calculating the number of correct responses. (d) *Picture completion (WAIS-R):* In this test (Tews, 1994), participants were asked to rearrange three series of pictures (comprising five or six each) such that the set of pictures will tell a coherent story. For this task, two scores were obtained (number of correct responses and task completion time).

*State anxiety.* To assess state anxiety, we employed the State-Trait Anxiety Inventory (STAI), comprising 40 items (Spielberger et al., 1983). This instrument aims to measure several dimensions of subjective strain in a more elaborate way, complementing the short measures of strain used as a manipulation check. In this study, we only employed the 20 items measuring state anxiety, with each of them using a 4-point Likert scale (i.e. total score can range from 20 to 80). We administered either the German version of the instrument (Laux et al., 1981) or the French one (Spielberger et al., 1983). This questionnaire was filled 3 times by the participants. One time at the very beginning of the experiment ($t_0$), one time after receiving feedback ($t_1$), and one time at the end of the experiment ($t_2$). Reliability of this scale in this study was satisfactory (McDonald’s omega, $\omega = 0.91$, 95% CI [0.88, 0.94]).[2]

*Emotion.* The Positive and Negative Affect Schedule (PANAS) was used to assess affect (Watson et al., 1988). Comprising 20 items, it makes use of a 5-point Likert scale, ranging from ‘very slightly’ or ‘not at all’ to ‘extremely’. To assess the emotional state of the participants, we administered a German version of the instrument (Breyer & Bluemke, 2016) or a French one, for which the items were translated from English (following the back-translation method) since the research literature did not offer a ready-to-use version. This questionnaire was filled in at the same time as the STAI. The reliability of the scale for positive affect was $\omega = 0.84$ (95% CI [0.78, 0.90]), and $\omega = 0.84$ (95% CI [0.78, 0.90]) for negative affect.

### 2.4. Procedure

Participants were randomly attributed to the three experimental conditions. The participants entered the laboratory in which they were welcomed by the experimenter. The experimenter gave the participants some instructions about the purpose of the study. Since the goal was to create social stress, it was required that a ‘cover story’ was provided (see below), which dissimulated the true nature of the experiment. Participants were informed that they might experience some stress during the experiment. After the oral instructions and the experimenter having responded to all the questions they might have had, participants were requested to read the form of informed consent carefully and to sign it

[2] Please note that due to issues related to the use of Cronbach’s alpha as a measure of internal consistency (see e.g. Dunn et al., 2014, for a summary), McDonald’s omega with 95% CI is reported in the present article.

afterwards.

As part of the cover story, participants were told that a new intelligence test for students was being developed by a university. The present study would help determine the qualities of the intelligence test. Participants were informed that they would have to complete a series of tests and several questionnaires. Having completed the first set of cognitive tests (cultural knowledge test, repeating numbers, numerical thinking; WAIS-R; Wechsler, 1981), the experimenter pretended to score the test. Participants were then given some bogus feedback about the test results (unless they were in the control condition with no feedback). If the feedback was negative, they were told that their IQ score was amongst the lowest ones of all student participants. This was demonstrated by using a large sheet of paper with a graph showing the test results. The experimenter added that the participants had not only had a very poor test score but had also been extremely slow in completing the test. Conversely, if the feedback was positive, they were told that their IQ score were amongst the best ones of all student participants. Again, a sheet of paper displaying the graph was used to underline the statement. The experimenter added that the participants had not only had a very high score but had completed the test extremely fast, too. In both feedback conditions, the provision of feedback was embedded in some discussion about the general purpose of intelligence testing. The choice of using an intelligence test to give feedback on was made to increase the strength of the manipulation in order to induce social stress. It was expected that results on such a personally and socially valued factor would be relevant to everyone and thus increase the impact of negative feedback.

Prior to the experimental manipulation in form of the cover story, participants completed a demographic questionnaire, followed by the completion of baseline assessments of PANAS and STAI ($t_0$). After the experimental manipulation, participants filled the PANAS and STAI again ($t_1$). Then, the following tests and questionnaires were completed: manipulation check, backward counting, d2, digit symbol coding, picture completion, and finally PANAS and STAI a third time ($t_2$).

Following the completion of the tasks and questionnaires, each participant was fully debriefed about the true nature of the study. First, the experimenter presented their apologies for providing incorrect information to the participant about the true nature of the experiment. The experimenter pointed out the need to misinform the participant to create the experimental conditions necessary for running the study. Furthermore, the experimenter pointed out the important applications of research of this kind, providing some examples of how this could help humans in the future (e.g., 'would there be a risk of negative feedback affecting subsequent performance of airline pilots?'). It was expected that this would increase the participant's understanding for the necessity to provide incorrect information as part of the experimental instruction. The participant was given the opportunity to ask questions about the experiment. Before the participant being paid and leaving the lab, the experimenter enquired whether the participant felt now at ease with the situation, following the debriefing. If the participant had still felt uneasy about the experiment, they would have been offered the possibility to make an immediate appointment at the in-house therapy center of the psychology department.

### 2.5. Data treatment

Following the experimental design, most measures were analyzed using a one-way analysis of variance, followed by Bonferroni's corrected pairwise comparisons for significant ANOVAs. If the homogeneity of variance and normality of distribution assumptions were both violated, a Kruskal-Wallis analysis of variance or a Wilcoxon rank-sum test was conducted. Additionally, one factorial analyses of covariance were conducted on variables measured before and after the experimental manipulation.

### 2.6. Results

#### 2.6.1. Manipulation check

For item 1, *"How do you evaluate your own performance on the test?*", participants rated their own performance differently depending on their condition; $F(2, 56) = 21.48$, $p = .001$, partial $\eta^2 = 0.434$. Participants in the positive feedback condition evaluated their own performance higher than the control group ($p < .001$) and the negative feedback group ($p < .001$). However, these two last groups did not differ significantly ($p = .29$). Item 2, *"How much stress are you feeling right now?*", showed no significant differences between conditions; $F(2, 56) = 0.748$, $p = .49$, partial $\eta^2 = 0.026$. For item 3, *"How stressful did you find the performance feedback?*", a *t*-test revealed significantly higher stress levels for participants having received negative feedback than those in the positive feedback condition; $t\,(36) = 2.68$, $p = .01$. Please note that this item was not administered in the control condition since it was not applicable. Overall, the statistical tests confirm the successful experimental manipulation of social stress through performance feedback, although one of the items was not significant.

#### 2.6.2. Performance

*Backward counting.* The number of errors in backward counting are presented in Table 1. The analysis of variance revealed no significant effect of feedback; $F(2, 59) = 2.673$, $p = .08$, partial $\eta^2 = 0.086$.

*Visual scanning. (d2-test).* Attentional performance showed no difference in both speed and accuracy subscales as a function of feedback (see Table 1). The analysis of variance confirmed this by revealing a non-significant effect for the speed score between conditions $F(2, 57) = 0.48$, $p = .62$, partial $\eta^2 = 0.017$. Similar results were found for accuracy score; $F(2, 57) = 1.80$, $p = .17$, partial $\eta^2 = 0.059$.

*Digit symbol coding.* In Table 1, the number of correct responses for the test involving digit symbol coding is presented. No significant effect of performance feedback was found; $F(2, 59) = 0.006$, $p = .99$, partial $\eta^2 = 0.000$.

*Picture completion.* The two performance scores for the picture completion test are shown in Table 1. The analysis showed a significant effect of feedback type for accuracy; $F(2, 57) = 3.459$, $p = .04$, partial $\eta^2 = 0.108$. A Bonferroni corrected post-hoc test showed that the scores for positive feedback were significantly higher than the scores for negative feedback ($p = .036$) but there were no significant differences between the two experimental groups and the control group. With regard to speed, the analysis revealed no significant effect of feedback type; $F(2, 57) = 0.983$, $p = .38$, partial $\eta^2 = 0.033$.

#### 2.6.3. Subjective measures

*State anxiety.* For this variable, the one factorial analysis of covariance at $t_1$, with $t_0$ as covariate, showed significant differences between conditions; $F(2, 56) = 5.34$, $p = .008$, partial $\eta^2 = 0.16$. Indeed, participants who received positive feedback had a lower state anxiety than the negative feedback group ($t(37) = 2.886$, $p = .006$) and the control group ($t(39) = -2.795$, $p = .007$). However, these two last groups did not differ significantly ($t(38) = 0.167$, $p = .868$), only partly supporting hypothesis 1a.

**Table 1**
Effects of type of performance feedback on the main dependent variables.

| Variable | Positive feedback Mean *(SD)* | Negative feedback Mean *(SD)* | No feedback Mean *(SD)* |
|---|---|---|---|
| *Manipulation checks* | | | |
| | *** | | |
| | | *** | |
| Item 1 | 7.37 *(1.92)* | 3.58 *(2.24)* | 4.57 *(1.32)* |
| Item 2 | 4.16 *(2.93)* | 5.05 *(2.37)* | 5.00 *(2.32)* |
| | ** | | |
| Item 3 | 3.37 *(2.41)* | 5.37 *(2.19)* | n.m. |
| *Performance* | | | |
| | * | | |
| Attention: accuracy (no. of items correctly marked) | 388.20 (34.14) | 367.37 (35.16) | 375.31 (34.95) |
| Attention: speed (no. of items worked through) | 428.00 (34.72) | 414.16 (54.27) | 417.43 (47.98) |
| Backward counting (no. of errors) | 1.35 (*.99*) | 2.16 (*2.01*) | 1.19 (*1.03*) |
| Digit symbol coding (no. of points) | 66.35 *(13.55)* | 66.26 (12.922) | 65.95 *(11.24)* |
| | * | | |
| Picture completion: accuracy (no. of points) | 1.37 *(0.34)* | 0.965 *(0.47)* | 1.11 *(0.60)* |
| Picture completion: speed (s) | 42.55 *(14.11)* | 40.70 *(13.45)* | 47.33 *(18.32)* |
| *Subjective measures* | | | |
| | ** | | |
| | | * | |
| Positive affect (10-50) | | | |
| $t_1$ | 31.80 *(6.33)* | 29.89 *(7.27)* | 32.10 (5.91) |
| $t_2$ | 31.00 *(6.48)* | 30.05 *(6.91)* | 32.19 (6.05) |
| Negative affect (10-50) | | | |
| $t_1$ | 16.85 *(5.33)* | 16.74 *(4.05)* | 18.33 *(5.95)* |
| $t_2$ | 17.0 *(5.12)* | 15.68 *(3.95)* | 17.38 *(5.10)* |
| | ** | | |
| | | ** | |
| State anxiety (20-80) | | | |
| $t_1$ | 36.85 *(9. 66)* | 40.00 *(11.24)* | 39.81 (9.84) |
| $t_2$ | 39.60 *(8.44)* | 37.32 *(7.91)* | 39.52 (8.00) |

*Notes:* * = $p < .05$; ** = $p < .01$; *** = $p < .001$; n.m. = not measured

*Positive affect.* A one-factorial analysis of covariance at $t_1$ with $t_0$ as covariate was calculated. The analysis revealed a significant main effect of type of feedback: $F(2, 56) = 6.264$, $p = .004$, partial $\eta^2 = 0.88$. Planned contrasts confirmed that there were significant differences between the condition positive and negative $t(37) = 3.458$, $p = .001$ whereby the subjects with the positive feedback showed a higher value in positive affect (see Table 1). The positive feedback group had lower positive affect than the control group, $t(39) = 2.428$, $p = .018$, while the control group and the negative feedback group did not differ significantly, $t(38) = 1.076$, $p = .287$. Hypothesis 1b for positive affect was only partly supported.

*Negative affect.* The one factorial analysis of covariance at $t_1$, with $t_0$ as covariate, did not show an effect of type of feedback on negative affect; $F(2, 56) = 1.886$, $p = .16$, partial $\eta^2 = 0.063$. Hypothesis 1b for negative affect was thus not supported.

### 2.7. Discussion

This is the first study that examined negative performance feedback using a wide range of established cognitive tests, allowing us to determine whether different types of cognitive performance would be vulnerable to this social stressor. Although participants felt stressed after having received negative feedback, their subsequent performance on the cognitive tests remained unimpaired on all tasks.

The primary outcome variables in this study were the different performance measures. Three out of the four measures were not affected by performance feedback, being in line with the 'blank out'-mechanism postulated by Sauer et al. (2019). Participants could protect their performance even though they were the target of social stress. Picture completion as the most creative task in the set of tasks showed higher accuracy levels in the positive feedback condition than for negative feedback whereas no such effect was observed for the speed component. This differential effect for speed and accuracy bears some similarity to the results of Alder's (2007) work, in which he found quality of performance to be affected by negative feedback but not quantity. This might also be explained in the framework of the speed-accuracy trade-off. A change in the speed-accuracy trade-off function under stress (i.e. involving a faster but less accurate response) has also sometimes been found when humans were exposed to stressors such as noise and time pressure (Hockey & Hamilton, 1983). An alternative explanation for this result is that the picture completion task, compared to the other tasks used, was the one most closely related to cognitive ability typically assessed by an intelligence test. It would then make sense that this task would be the most affected by negative feedback on a preceding intelligence test.

Although recent work modelling social stress found similar results, with performance on four different tasks being unimpaired (Peifer et al., 2020), there is overall an inconsistent results pattern in the very small number of studies examining social stressors and objective performance. There is also empirical work that was in line with the predictions of the 'rumination'-mechanism (i.e. performance decrease; Lustenberger & Jagacinski, 2010) or the 'increased-motivation'-mechanism (i.e. performance increase; Byrne et al., 2016). The support for the 'blank out'-mechanism found in the present study and some other work may be considered a positive finding. Indeed, it may carry the practical implication that performance levels may be maintained by operators (especially in safety-critical jobs) despite negative effects at the subjective level.

The results for the subjective measures showed a different pattern. The manipulation checks indicated overall that participants were negatively affected by negative feedback compared to positive feedback. The results for positive affect revealed significantly higher ratings for positive feedback than negative feedback and the control group. Our first hypothesis was only partly confirmed in that positive feedback induced less state anxiety compared to receiving negative feedback or not receiving any feedback at all. This is an interesting finding, which may suggest that even if participants were subject to a stressful negative feedback (as shown by the manipulation check), this did not spread over to the general state of anxiety. This might indicate that specific or stimulus-related stress states do not necessarily affect general states of stress. As a last point, we would like to add that a considerable number of participants, after having received negative feedback, attempted to offer excuses to the experimenters to justify their poor performance. For example, some participants said they would have performed better in different types of tasks, or that they were tired and they did not sleep well the previous night. This may also be taken as an indication of the successful experimental implementation of negative feedback as a stressor since such justifications might also be observed in a work context.

## 3. Study 2

### 3.1. Goal of the study

The main goal of the second study was to examine whether negative feedback coming from a computer produces different effects than when coming from a human. This was implemented by using a modern form of computer feedback, and focusing on the source of feedback while keeping the medium constant.

The second goal was to investigate whether performance on a different set of tasks would be affected differently, using a stronger form of negative feedback, which contains elements of destructive feedback.

### 3.2. Participants and experimental design

A total of 89 students (50.5% female), aged 18–35 years ($M = 22.48$; $SD = 2.84$), participated in the study, all of which were French native speakers. They were recruited from the different faculties of the University of Fribourg and schools of higher education, with the exception of students from psychology and related sciences (e.g., education). They were not allowed to take part in the experiment since they may be familiar with experimental scenarios using deception. Participants received CHF 20.- as a financial compensation for their participation.

A one-way between-subjects design was used in this experiment. The independent variable 'feedback source' was manipulated at three levels: human source, computer source or no feedback (i.e. control group).

### 3.3. Dependent variables

#### 3.3.1. Manipulation checks

Several measures were used for the manipulation check. First, an item was created to verify whether the induction of stress was successful between the two experimental groups and the control group: "*To what extent are you feeling stressed?*" (with a 7-point scale ranging from *not at all* to *a great deal)*. Second, another item measured subjective state anger on a scale ranging from 1 (*not at all*) to 7 (*a great deal*): "*To what extent are you feeling angry?*". This was based on Baron (1988), who showed that destructive feedback induces anger. All manipulation checks were administered twice: at the very beginning of the experiment and right after participants received feedback.

#### 3.3.2. Performance

*3.3.2.1. Attention and concentration: d2-R (speed and accuracy).* Attentional performance was measured by using the d2-R (Brickenkamp, 1962). Representing a sustained visual scanning task, this test allowed us to assess both quantitative and qualitative aspects of performance. The speed score was calculated by the number of items worked through, whereas for the accuracy score the number of errors made was subtracted from the total number of items worked through. While several test scores can be used in the d2-test, only the scores for speed and accuracy are considered to meet psychometric criteria for reliability (Steinborn et al., 2018).

*3.3.2.2. Convergent creativity: Remote Associates Task.* Convergent creativity was measured by an adapted version of the Remote Associates Task (Mednick, 1968). This test consists of sets of three words, and the goal is to find a new word that is related to the three words presented in the test item. The new word can be a synonym, semantic association or compound word. For example, the answer for the item "home – sea – stomach" is the word "sick", as people can be homesick, seasick or sick in the stomach.

As there is currently no French version of the Remote Associates Task, items were specifically developed for this study (see section 3.4). 15 items were chosen and balanced out for difficulty: 5 difficult items (17–26% correct answers from pilot study), 5 moderately difficult items (50–54%), and 5 easy items (76–84%). The performance score used in the experiment was the total number of correct responses to the 15 test items.

*3.3.2.3. Divergent creativity: Alternate Uses Task.* Divergent creativity was assessed by the Alternate Uses Task (Guilford, 1960). In this test, participants were asked to list the potential uses of a brick. The instructions were formulated such to encourage participants to find truly original and creative answers (Runco et al., 2005). Performance was measured by the number of valid answers (fluency score) and the degree of originality. The scoring procedure for determining degree of originality was taken from O'Connor et al. (2013). Answers given by less than 1% of the participants were scored 2 points, and answers given by less than 5% of participants were scored 1 point. To control for higher originality being due to higher fluency, an index was calculated (originality/fluency).

#### 3.3.3. Subjective measures

*Affect.* The Self-Assessment Manikin scale (Bradley and Lang, 1994) was used to assess the affective state of participants on two dimensions: valence (negative vs positive affect) and arousal (low vs high). This 9-point scale was administered twice: once at the very beginning of the experiment and once right after participants received feedback.

*Interpersonal fairness.* Four items were selected from the interpersonal fairness subscale of the Organizational Justice Scale (Colquitt et al., 2015) to assess the fairness of the feedback source for the participants. The items were translated into French (using the back-translation method) and slightly adapted to the experiment. The four items (and the instructions preceding them) were worded as follows: "*The questions below refer to the person/program who formulated the feedback you received at the beginning of the experiment. (a) To what extent did she/he/it treat you with dignity? (b) To what extent did she/he/it treat you in a polite manner? (c) To what extent did she/he/it treat you with respect? (d) To what extent did she/he/it refrain from improper remarks or comments?*" All items were rated on a 5-point Likert scale (ranging from *not at all* to *a great deal*). The scores from all four items were averaged to obtain an overall fairness score. Reliability of this scale in this study was satisfactory (McDonald's omega, ω = 0.94, 95% CI [0.92, 0.97]).

*Level of distraction.* Six 5-point scale items (ranging from *strongly disagree* to *strongly agree*) were used to measure the level of feedback-related distraction from the task after receiving feedback. Two items were taken from Alder's scale (2007) and translated into French, using the back-translation method (i.e. "*The feedback I received helped me focus my attention on the task*" (reverse scoring), and "*The feedback I received was often a distraction*"). The following four items were specifically developed for this study, based on the Feedback Intervention Theory (Kluger & DeNisi, 1996): "*For the rest of the experiment, I often thought about the feedback I received*", "*I often felt threatened by the feedback I received*", "*The feedback I received made me question my abilities*" and "*I was annoyed by the feedback I received*". The scores of all items were averaged to obtain a global score of task-unrelated attention, with a high score indicating that attention was focused on feedback and the self rather than the task. The reliability of the scale was ω = 0.81, 95% CI [0.74, 0.89].

*Desire to improve.* To measure the participants' desire to improve their performance after feedback, two items from Alder (2007) were translated into French, employing the back-translation method ("*I felt I wanted to improve my performance in response to the feedback I received*", "*I tried to work harder after I had received feedback on my performance*"). A 5-point scale ranging from '*strongly disagree*' to '*strongly agree*' was used. The scores from both items were averaged. Reliability was satisfactory, Spearman-Brown = 0.84.

*State self-esteem.* The State Self-Esteem Scale (SSES; Heatherton & Polivy, 1991) was administered after the last task of the experiment. It consists of 20 items using a 5-point scale (ranging from *not at all* to *extremely*), and three sub-scales: self-esteem related to performance (7 items; ω = 0.83, 95% CI [0.78, 0.88]), appearance (6 items; ω = 0.83, 95% CI [0.77, 0.88]) and social (7 items; ω = 0.86, 95% CI [0.82, 0.90]). For the purpose of this experiment, the overall score was calculated by aggregating the scores of all 20 items.

### 3.4. Pilot studies

Two pilot studies were conducted before starting Study 2. The first one was to determine the appropriateness of the wording of the negative feedback used to verify its destructive nature. The second pilot study was necessary to create French items for the Remote Associates Task, since no established French version is available.

#### 3.4.1. Pilot study on feedback destructiveness

In Study 2, negative feedback was given in a destructive manner, following Baron's (1988) principles. According to Baron, destructive feedback is too general, inconsiderate, contains threats and attributes poor performance to internal factors. This meant that feedback contained expressions such as "*Extremely low score*" or "*Seems to have had great difficulties with a rather simple task*". The pilot study (N = 15) was conducted to ensure that the destructive feedback used in the main study increased the strength of the experimental manipulation. The wording of the feedback was presented to participants in text form, with the following questions asked (using a 7-point Likert scale): (item a) "In your opinion, was the feedback presented in a rather sensitive or insensitive way?" (scale ranging from '*very sensitive*', 1, to '*very insensitive*', 7); (item b) "In your opinion, did the presented feedback contain threats?" (scale ranging from '*no threats*', 1, to '*containing threats',* 7); (item c) "In your opinion, was the feedback presented rather specific or general in content?" (scale ranging from '*very specific*', 1, to '*very general*', 7); and (item d) "In your opinion, did the feedback attribute the performance to causes that are external or internal to the participant?" (scale ranging from '*internal causes',* 1, to '*external causes*', 7). Overall, the destructive nature of the feedback was confirmed: feedback was perceived to be very insensitive (item a; $M = 6.2$, $SD = 0.91$), general rather than specific (item b; $M = 6.27$, $SD = 1.06$), and performance was attributed to internal causes (item c; $M = 1.53$, $SD = 0.81$). However, the feedback was not judged as containing threats (item d; $M = 2.8$, $SD = 1.64$).

#### 3.4.2. Remote Associates Task items

For the Remote Associates Task, sixty-nine items were created in French, based on the original items of the English version (Thuillard & Richter, unpublished). Item difficulty was assessed using an online questionnaire (N = 187). The complete set of 69 items was divided into three lists of 23 items. Before completing the online questionnaire, participants were randomly assigned to one of the three lists. The items on the assigned list were presented one by one, with a time limit of 1 min per item. On average, each item was tested by 62 participants. This allowed us to determine item difficulty based on the percentage of participants who found the correct answer (0–100%), and select the items to use in the study based on this percentage (see section 3.3.2).

### 3.5. Procedure

Participants were recruited by email, which was sent out to students from selected university faculties (see section 3.2). When students accepted to take part in the study, they were invited to the laboratory. Having arrived at the laboratory, they received an information sheet providing a cover story dissimulating the real purpose of the study, and a consent form that they were asked to sign before starting the experiment. The cover story was that they would have to perform several attention and creativity tasks with a view to investigating the link between attention and creativity. Participants were randomly assigned to conditions in the following way: each participant code was assigned to a particular condition beforehand, then each person taking part in the experiment received their code and condition simultaneously. In the human source condition, participants were informed that they would receive performance feedback on the first task from the supervisor of the experimenter (the supervisor was not visible to the participant). Giving performance feedback was justified by explaining that it represents a common procedure in experimental psychology because it improves data quality and ensures that participants feel more involved during the course of the experiment. Participants in the computer source condition received the same cover story, except that a newly developed deep-learning-based software rather than a human would provide their feedback. Participants in the control group did not receive any feedback. Feedback was given on the participants' performance on the first task, which was a difficult version of the Remote Associates Task based on recommendations from McFarlin and Blascovich (1984). We chose 10 extremely difficult items (less than 10% of correct answers in pilot study), three relatively difficult items (about 30%), one item of medium difficulty (50%), and an easy item (88%). This manipulation allowed us to decrease the performance of participants, which made the faked negative feedback subsequently given to participants more credible.

*Human feedback.* The manipulation in the human source condition was performed as follows: after having completed the first task, participants had to wait for 5 min while their performance data was corrected and analyzed by the supervisor of the experimenter (according to the cover story). After 5 min, the experimenter went to fetch the handwritten feedback from the supervisor. The feedback sheet was placed inside an opaque folder and afterwards handed out to the participant. Participants were informed that the experimenter was not allowed to read it and did not know its content in order not to influence the results of the experiment. This prevented participants from justifying their poor performance to the experimenter. After receiving feedback, the participants were left alone to read it, and then continued the experiment as soon as they were finished reading.

*Computer feedback.* In the computer source condition, a purpose-built, fake automatic correction software was presented on the participant's computer screen to increase the credibility of the manipulation.

The software pretended to load and analyze the data from the first task, finally printing out the feedback on paper in another room. Following the same procedure as in the human feedback condition, the experimenter went to fetch the printed feedback and handed it over to the participant in the same opaque folder. It was programmed to take 5 min in order to match the waiting time in the other conditions. In this way, feedback was provided to the participant in a similar form as in the human source condition, that is, matching feedback medium used in both conditions. The information sheet printed out in the computer feedback condition was the same as in the human feedback condition, except for a computer printout being used instead of handwriting. Thus, only the feedback source differed from the human feedback condition.

*No feedback condition.* In the control condition, participants were simply told that they would have a 5-min break after the first task. They received the same cover story as in the other conditions.

*Remainder of experiment and debriefing.* After the experimental manipulation, all participants completed the remaining part of the experiment, which did not differ between conditions. Performance tests were completed in the following order: d2-test (lasting 5 min), Remote Associates Task (maximum time of 15 min) and Alternate Uses Task (maximum time of 5.5 min). Before the end of the experiment, participants completed the state questionnaires in the following order: state self-esteem scale, interpersonal fairness, level of distraction, and desire to improve. After the experiment, participants were debriefed in the same way as in study 1.

### 3.6. Data analysis

The data for each dependent variable were analyzed in the same way as in study 1 (see section 2.5). Additionally, hypothesis 2b required a different procedure since it predicted a nil effect of negative feedback on attentional performance. Based on Cortina and Folger (1998) and Onnasch (2015), we adapted alpha to a 20% level for the relevant analyses.

### 3.7. Results

#### 3.7.1. Manipulation checks

The item (‘*To what extent are you feeling stressed?*‘) being used as a manipulation check showed that participants perceived general stress to be higher in the two experimental conditions (see Table 3) than in the control group. The one-factorial analysis of covariance, with pre-feedback stress as covariate, proved this difference to be statistically significant; $F(2, 85) = 9.91$, $p < .001$, partial $\eta^2 = 0.117$. Bonferroni-corrected post-hoc comparisons showed a significant difference between human condition and control ($p < .001$), and marginally non-significant between computer condition and control ($p = .052$). The two feedback conditions did not differ significantly ($p = .24$). Post-feedback anger differed between conditions; $F(2, 85) = 11.14$, $p < .001$, partial $\eta^2 = 0.208$, with pre-feedback state anger being used as a covariate. It was rated significantly higher in both the human ($p < .001$) and the computer source ($p = .002$) conditions than in the control group. However, the human and computer conditions did not differ significantly ($p = 1.0$). Overall, these results indicate that the experimental manipulation of feedback as a source of stress was effective.

#### 3.7.2. Performance

*Convergent creativity performance (Remote Associates Task)*. The scores for convergent creativity performance are presented in Table 2. No significant differences between conditions were found in the Remote Associates Task; $F(2, 86) = 1.67$, $p = .19$, partial $\eta^2 = 0.037$.

*Divergent creativity performance (Alternate Uses Task)*. The data for both aspects of divergent creativity (i.e. fluency and originality) are shown in Table 2. The fluency scores differed as a function of experimental conditions; $F(2, 86) = 3.51$, $p = .03$, partial $\eta^2 = 0.075$. Post-hoc analyses with Bonferroni correction showed that the fluency score in the computer condition was significantly higher than in the control group; $p = .047$. No other post-hoc comparison was found to be significant. For response originality, no significant effect was observed; $F(2, 86) = 0.74$, $p = .48$, partial $\eta^2 = 0.017$. To test for a possible shift in speed-accuracy trade-off, we also tested the index as a ratio of fluency and originality but found no difference between experimental conditions; $F(2, 86) = 0.65$, $p = .52$, partial $\eta^2 = 0.015$. Overall, hypothesis 2a was not supported.

**Table 2**
Means and standard deviations of performance as a function of feedback source.

| Variable | Human feedback Mean *(SD)* | Computer feedback Mean *(SD)* | No feedback Mean *(SD)* |
|---|---|---|---|
| Attention: accuracy (no. of items correctly marked) | 513.6 *(65.3)* | 526.7 *(85.5)* | 521.4 *(55.1)* |
| Attention: speed (no. of items worked through) | 532.5 *(69.5)* | 546.7 *(85.9)* | 541.13 *(57.0)* |
| Convergent creativity (0–15) | 9.1 *(2.9)* | 8.52 *(3.0)* | 7.77 *(2.6)* |
| Divergent creativity: fluency (no. of points) | 4.41 *(2.3)* | 6.0 *(3.6)* * | 4.16 *(2.6)* * |
| Divergent creativity: originality (no. of points) | 1.55 *(1.6)* | 2.14 *(2.3)* | 1.84 *(1.5)* |

*Notes:* * = $p < .05$.

*Attentional performance (d2-test)*. The data for attentional performance in the form of accuracy and speed are presented in Table 2. For accuracy, the analysis of variance revealed no significant difference between conditions; $F(2, 85) = 0.26$, $p = .77$, partial $\eta^2 = 0.006$. Similar results were found for the speed performance score, with the analysis of variance showing no significant difference between human feedback, computer feedback and the control group; $F(2, 85) = 0.29$, $p = .75$, partial $\eta^2 = 0.007$. These results provide additional evidence for acceptance of hypothesis 2b suggesting a nil effect.

#### 3.7.3. Subjective measures

*Affect.* Table 3 shows the valence scores of assessing affect by means of the Self-Assessment Manikin. The one-factorial analysis of covariance, with pre-feedback valence as a covariate, showed that after receiving feedback affect scores differed between conditions; $F(2, 85) = 17.53$, $p < .001$, partial $\eta^2 = 0.292$. Bonferroni-corrected post-hoc comparisons showed that valence was rated significantly higher in both the human and the computer source conditions than in the control group (both comparisons: $p < .001$). However, the human and computer source conditions did not differ significantly from one another ($p = .23$). No significant differences were found for arousal; $F(2, 85) = 2.33$, $p = .10$, partial $\eta^2 = 0.052$. Hypothesis 2c was supported by these results.

*State Self-Esteem Scale*. The data in Table 3 show that there was no difference in self-esteem between the three conditions when examining the total score by using analysis of variance; $F(2, 86) = 0.82$, $p = .44$, partial $\eta^2 = 0.02$. The most relevant subscale ‘performance’ showed a marginally significant difference, with the score in the two experimental groups being lower than in the control group; $F(2, 86) = 2.85$, $p = .06$, partial $\eta^2 = 0.06$. These results do not support hypothesis 2d.

*Perceived fairness.* Perceived fairness was rated significantly higher in the human feedback condition than in the computer feedback condition (see Table 3). The Wilcoxon rank sum test confirmed this difference to be significant; $W = 583$, $p = .007$.

*Level of distraction.* As the data in Table 3 show, the level of distraction was significantly higher in the human source condition than in the computer source condition; $W = 534$, $p = .04$. This indicates that the attention of participants was more strongly focused on the task in the computer source condition than in the human source condition.

*Desire to improve.* The data for the variable ‘desire to improve’ are

**Table 3**
Means and standard deviations of manipulation checks and subjective variables as a function of feedback source.

| Variable | Human feedback Mean *(SD)* | Computer feedback Mean *(SD)* | No feedback Mean *(SD)* |
|---|---|---|---|
| *Manipulation checks* | | | |
| Perceived stress (1-7) | 2.97 *(1. 68)* | 2.93 *(1.25)* | 2.13 *(1.34)* |
| State anger (1-7) | 2.93 *(1.87)* | 2.45 *(1. 5)* | 1.39 *(0.72)* |
| *Subjective measures* | | | |
| Affects: valence | 4.95 *(2.29)* | 5.28 *(1.46)* | 6.84 *(1.34)* |
| Affects: arousal | 4.86 *(1.64)* | 5.00 *(1.49)* | 4.19 *(1.25)* |
| Interpersonal fairness (1-5) | 2.0 *(1.18)* | 1.4 *(.82)* | n.m. |
| Level of distraction (1-5) | 3.18 *(1.0)* | 2.67 *(0.76)* | n.m. |
| Desire to improve (1-5) | 3.71 *(1.24)* | 3.33 *(1.31)* | n.m. |
| State self-esteem (20-100) | 71.0 *(12.99)* | 72.96 *(13.22)* | 75.19 *(11.92)* |
| Performance self-esteem | 25.38 (*5.71*) | 26.93 (*4.61*) | 28.26 (*3.46*) |

*Notes:* * = $p < .05$; ** = $p < .01$; *** = $p < .001$; n.m. = not measured

presented in Table 2. The analysis showed that the desire to improve performance after feedback did not differ significantly between the human source condition and the computer source condition; $W = 476.5$, $p = .26$.

### 3.8. Discussion

The main goal of the second study was to examine whether performance would be impaired after using a destructive form of negative human feedback, using a different set of tasks, and employing a computer as an additional source of feedback. The replication was successful for the performance measures (i.e. attention and creativity), with negative feedback from a human having no negative influence on subsequent performance. The same results emerged when negative feedback was provided by a computer, with one exception in an unexpected direction (i.e. a small effect of negative computer feedback increasing fluency in divergent creativity). Overall, the findings for performance are again in support of the 'blank out'-mechanism.

The analysis of the subjective variables (which were larger in number than in the first study) indicated that performance protection was paralleled by considerable changes in the participants' subjective state. Negative feedback increased stress, anger and negative mood in both experimental groups, which is consistent with previous research on negative human feedback (e.g., Baron, 1988; Nummenmaa & Niemi, 2004; Raver et al., 2012). The observation that negative computer feedback induced the same affective reactions as human feedback represents an important result considering the lack of recent empirical data on computer feedback and affect. Both experimental groups reported being motivated to improve their performance after having received negative feedback, though human feedback was felt to be more distracting from the task than computer feedback. Surprisingly, the participants who received negative feedback were not affected in their state self-esteem, which seems to be opposed to what could be expected based on the 'Stress-as Offense-to-Self'-approach (Semmer et al., 2019). Interestingly, the evaluation of interpersonal fairness differed between the two experimental groups. Negative computer feedback was perceived as being more unfair than the same feedback coming from a human, indicating an effect of feedback source.

Overall, the subjective measures seem to indicate that some costs are associated with the successful protection of task performance under social stress no matter whether a human or a computer is responsible for these stressful conditions. Additionally, automated negative feedback does not appear to be a like-for-like replacement of human feedback due to its perceived unfairness, which calls for some caution when considering its implementation in the workplace.

## 4. General discussion

The main goal of this article can be summarized in three points: (a) to examine whether negative performance feedback as a social stressor affects subsequent performance, (b) whether different types of tasks were affected to a different extent, and (c) to determine whether the effects were different when feedback was provided by a computer rather than a human. Overall, the two studies indicated that negative feedback did not impair performance, and this regardless of feedback source or task type, though some subtle effects emerged. It appeared that performance maintenance under stress came at a cost in form of stress-related effects being detected at the subjective level. Finally, negative human and computer feedback were found to be perceived differently by recipients on several dependent variables.

### 4.1. Performance

In both studies, the manipulation check indicated a successful induction of social stress, and in both studies, no effects of negative human feedback on performance were found even though a wide range of tasks was used. This pattern of results supports the 'blank out'-mechanism, according to which participants are capable of focusing on the task such that they are impervious to the effects of social stress at the performance level. The research literature did not provide consistent support for the 'blank out'-mechanism, with negative feedback sometimes impairing subsequent task performance (e.g., Alder, 2007; Alder & Ambrose, 2005; Nease et al., 1999; Raver et al., 2012), as predicted by the 'rumination'-mechanism. However, the present findings are in line with the results of some recent studies, which found that participants were able to protect their performance from the stress caused by inadequate human feedback (Peifer et al., 2020), or a combination of negative feedback and social exclusion (Sauer et al., 2020). Together, studies 1 and 2 measured the following types of performance: backward counting, attention, picture completion, symbol coding and convergent and divergent creativity. Although this wide range of tasks covering a large spectrum of cognitive abilities was used in all this work, performance was never impaired following negative feedback. The only significant effect on performance detected, an improved performance on picture completion, appeared subsequently to positive human feedback. These results also do not appear to support Van Dijk and Kluger's findings (2011) that negative feedback affects performance differently depending on task type. Alternatively, it may be that well-established cognitive tests, such as used in study 1, represent a type of task that is unaffected by negative human feedback. Overall, the main implication of these results is that it appears possible for humans to protect their performance from the impact of social stressors, though this might be associated with some cost.

While performance was largely protected from social stress in the form of negative performance feedback, there were several indications of this stressor having an impact at the subjective level. The manipulation checks provided a first sign of the presence of social stress in participants. Then, we observed a negative impact in both studies on affective measures (with slight variations in the type of variable affected) but also on self-reported task management behavior (i.e. task-related distraction). This pattern may suggest that there are some costs associated with protecting task performance under social stress, which are in line with the model of compensatory control mechanism (Hockey, 1997). This model predicts active human performance management to protect overall task performance, though this protection may have costs at the cognitive-energetical level. In the present study, the costs observed were of a slightly different nature. In the two experimental conditions, less positive affect was observed compared to the no-feedback condition. This may also indicate some evaluative process to digest the negative feedback even if the hypothesized effect for self-esteem was rather small (and just not reached the level of significance). However, it must be noted that this protection of performance was detected as negative feedback was given only once. The subjective costs associated with performance protection might make it more difficult to keep maintaining performance in case of repeated social stress induction or in the long term.

### 4.2. Computer feedback

With regard to one of our main research questions addressing the effects of feedback source on performance, we found that computer and human feedback both showed almost no effects. Only one minor difference emerged in that participants who received negative feedback from a computer found more uses in the Alternate Uses Task than the control group. No such difference was found for participants who received human feedback. While it may represent a case of 'increased motivation'-mechanism (as described in Sauer et al., 2019), it would be surprising to observe this mechanism on only one subscale and not on the other performance measures. Despite this minor difference being observed, there seemed to be overall little impact of feedback source on performance. However, the overall pattern shows some differences compared to the reference study by Van Dijk and Kluger (2011), which may be related to the different set-up of the two studies. First, in Van Dijk and Kluger's study a computer did not only generate the feedback (i.e. source) but also delivered it (i.e. medium) whereas in the present study, a computer generated the feedback but a human delivered it. Second, in contrast to the other study, we used in our experimental cover story the concept of deep learning, with a view to investigating a modern form of automation as well as increasing the credibility of the computer feedback. Both aspects may have worked in the same direction, contributing to a higher similarity of the two feedback types. This is because computer generated feedback based on deep learning may make computer feedback appear more similar to human feedback than when employees assume that a simpler form of computing had been used. Indeed, since feedback content was exactly the same in both experimental conditions, the text was analogous to what a human could write. At the same time, the two feedback conditions become more equivalent if they both use a human to deliver the feedback (as in the present study). While keeping the feedback medium stable and varying only the feedback source (as in the current work) enjoys the advantage of being able to isolate the effect of source, it has the disadvantage of reducing the distinctiveness of the two types of feedback (i.e. computer and human feedback become more similar). This may have led to a lower probability of demonstrating a distinct difference in performance between human and computer feedback.

### 4.3. Subjective variables

Participants who received negative feedback (either from a human or from a computer) were stressed, angry and in a negative mood. This might suggest that maintaining performance following computer feedback would be associated with some cost comparable to human feedback. Thus, similar concerns could be raised about the effects of repeated negative computer feedback over a longer period of time. However, the pattern of results for human and computer feedback is not completely the same. Although both experimental groups reported being equally motivated to improve their performance after feedback, this motivation resulted in improved performance only in one task for the computer group. This difference might be explained by the fact that participants reported being more focused on the task when receiving computer feedback than human feedback. This would represent a positive aspect of computer feedback being less distracting than human feedback. As shown by Raver et al. (2012), in case of destructive human feedback participants are more likely to blame and distrust the feedback-giver while also thinking he or she intended to harm them. This negative reaction might be attenuated if the feedback-giver is a computer.

In contrast to what was expected, state self-esteem was not affected in either of the two experimental groups, which is quite surprising considering the effects detected for stress and affect. According to the SOS approach (Semmer et al., 2019), negative feedback can induce stress by thwarting important goals (such as maintaining a positive self-view), or through insufficiency or disrespect. These three mechanisms should all be able to unfold in case of destructive negative feedback. At first view, this does not appear to be in support of this aspect of the SOS approach. Indeed, threats to the self are a core postulate of this model, and effects of social stress on self-esteem have been found in the literature (e.g., Eatough et al., 2016; Schulte-Braucks et al., 2019). Yet, there are several possible reasons that could explain why self-esteem was not affected. As explained previously, there was only one induction of social stress through negative feedback. Even though a state self-esteem scale was used, which is sensitive to quick changes, it is possible that effects on self-esteem may be more easily detected following repeated, longer or more intense stress inductions. However, the SOS approach also hypothesizes protection mechanisms, which could explain that no effects on self-esteem were found. For example, it is possible to protect oneself from negative feedback by attributing it to a lack of fairness of the feedback source (Semmer et al., 2019). Crucially, this effect on fairness has been found in our study. Believing that the feedback was unfair may have helped the participants to protect their self-esteem, which would have otherwise been threatened by negative feedback.

Perhaps the most remarkable effect of feedback source in the present work is that negative computer feedback was considered as less fair than negative human feedback. This means that although feedback was provided in exactly the same way and had the exact same content in both conditions, participants felt they were treated more unfairly by the computer than by the human. This difference is in line with previous work, which found face-to-face feedback to influence interpersonal fairness (Alder, 2007; Alder & Ambrose, 2005) because it gives recipients a possibility to offer excuses for their poor performance. For example, such behavior was observed by the experimenters in Study 1 when providing negative performance feedback to participants on a face-to-face basis. Providing justifications for poor performance may help participants to cope better with negative feedback, which hence would increase perceived fairness. However, it is much more difficult for employees to offer excuses to an automated system by providing explanations for their unsatisfactory performance. Crucially, our work goes even beyond these findings in that it shows that differences in interpersonal fairness arise even when human feedback is not given face-to-face, indicating that regardless of how feedback is delivered, there might be something in automated feedback that is perceived as fundamentally more unfair. This result could be interpreted within the SOS mechanism of unfairness being attributed to feedback source as a means to protect self-esteem. Since computer feedback was felt as more unfair than human feedback, it could mean that it posed a greater threat to the participants' self-esteem, and thus required a stronger attribution of unfairness in order for self-esteem to still be protected. Alternatively, this result could be seen as a possible violation of etiquette in human-machine communication. Since in Study 2 human and computer feedback had exactly the same content, computer feedback included characteristics of human communication. While this was congruent with the cover story of a "deep-learning based automated system", it might at the same time have violated participants' expectations of how a machine should communicate with them.

This difference in fairness perception may raise some concerns about the consequences of the increasing tendency to provide automated feedback. In the literature, interpersonal fairness has been linked to task performance, organizational citizenship behavior and counterproductive work behavior (Colquitt et al., 2013), which could all be impaired by automated feedback that is not perceived as fair. While the idea of interpersonal interactions between a human and a system could be counterintuitive, this was already proposed by Alder (2007). Though he found no differences in interpersonal fairness between human and computer feedback, our result could be a sign that, since his study, perceptions and attributions in society regarding automation have changed. Indeed, computers using complex algorithms are playing an increasingly important role in work environments, even entering the leadership domain (Lee et al., 2015; Wesche & Sonderegger, 2019), in which negative performance feedback is part of leader-follower communication. Lower perceived interpersonal fairness could be a sign of lower acceptance of automated agents and of their decisions at work in general. Interestingly, another study recently showed that in decision situations, automated agents have the same impact on procedural fairness as humans (Ötting & Maier, 2018). Although an experimental setting may make it more difficult to distinguish between these different facets of fairness, future research should consider these different aspects when investigating human-machine interaction in work domains.

### 4.4. Generalization of results in work settings

It is possible that negative feedback would have different effects in real work settings when given to employees rather than to students. In the case of employees working in safety-critical settings, it could be imagined that they would be able to protect their performance as well, due to the pressure of preventing highly undesirable consequences. However, in less critical settings, it is also possible that employees would be more affected by negative feedback. Indeed, people usually see their professional role as part of their identity (Ashforth & Schinoff, 2016). Therefore, and according to the SOS approach (Semmer et al., 2019), negative performance feedback should represent a bigger threat to the self to an employee than to a student. It might then be more likely for subsequent performance to decrease as a result of the 'rumination' mechanism, as postulated in Sauer et al. (2019). Even in the case where performance could be maintained regardless, we could imagine that greater subjective costs would be associated to this protection. It must be noted however that since the cognitive tasks used in the present studies are not closely modelled on real work activities (see Limitations section below), our results cannot be directly transferred to actual work settings.

### 4.5. Limitations and future studies

There are several limitations of the studies. (a) Most analyses were conducted in an exploratory way. This means that the hypotheses tests should be interpreted with some caution since power may have been insufficient to detect some other effects. This limitation mostly applies to the performance measures. After Study 1, we used in Study 2 bigger samples and a stronger manipulation, and still found no effect of feedback on performance. Despite these results, it is still possible that there are effects that we could not detect. Therefore, our results will need to be replicated in future research.

(b) The current studies modelled preceding exposure to the social stressor (i.e. performing after having been given negative performance feedback) rather than simultaneous exposure (i.e. performing while being given negative performance feedback). Both represent typical situations at work. However, we assume that intermittent (i.e. repeated) or continuous exposure to the social stressor may represent a higher intensity of the stressor than post-exposure task completion. Therefore, we would recommend that future research examines the impact of intermittent or continuous exposure to the social stressor to determine whether the 'blank out'-mechanism would also be observed under these aggravated conditions, and whether the costs of maintaining performance would be the same. (c) We used (static) cognitive tests rather than cognitive tasks that were closely modelled on real work activities. On the one hand, this may have reduced ecological validity. On the other hand, such tests have good psychometric properties (high objectivity, high reliability, and high predictive validity), being very widely used in personnel assessment to predict future job performance. It is difficult to assess the impact of this limitation on our findings. Considering how task type appears to moderate how performance is affected by social stress, future studies could also test more tasks focusing on specific cognitive functions that have not been used yet. Additionally, given that the present work used a series of non-dynamic single tasks, future work should envisage making more use of tasks modelling real work activities. This may also include the use of multiple-task environments that allow a distinction to be made between primary and secondary tasks. All the cognitive tests used were equivalent to primary tasks (even if they tested low-level cognitive abilities like visual scanning), which required no trade-off in allocating cognitive resources to different tasks. Considering our results, future research should continue to use subjective measures to complement objective performance measures. (d) Although precautions were taken to avoid them, we cannot completely rule out experimenter demand effects. Such effects are less likely to appear for performance measures, since the goal of these tasks is to perform as well as possible. However, it is possible that subjective measures such as fairness, state stress or state self-esteem raised suspicions and cued participants in the experimental conditions that we expected them to be in a more negative state than at the beginning of the experiment. This is why we decided to use a sample of non-psychology students since they usually are less familiar with experimental manipulations.

Fairness appears to be a promising concept to explain the underlying processes, in particular in the context of automation. In addition to interpersonal fairness, other types of fairness could be measured with regard to the effects of social stress in human-machine teams. Future studies on fairness could try to understand what exactly in automated feedback is more unfair than human feedback. In this regard, it would be interesting to investigate human attribution when they receive negative feedback from some form of automation. Indeed, in this situation, the human could always blame other humans for the automated feedback. For example, they could blame the people who programmed the machine, or the experimenter for deciding to use it. We could also imagine that attribution could vary depending on how "modern" the automation is perceived. The more modern and autonomous the automated system is perceived, the less likely it might be to attribute the blame to humans. Future research could thus investigate these attribution processes with different levels of automation, ranging from low (e.g. an old computer) to high technological level (e.g. AIs). Alternatively, future studies could investigate whether fairness is linked to machine etiquette. Etiquette in human-machine interaction can influence perception of automation, such as trust (Parasuraman & Miller, 2004). Poor machine etiquette, such as impoliteness, might possibly impact perceived interpersonal fairness. It could once again be interesting to use more or less "modern" automated systems, since humans might have different etiquette expectations depending on how advanced the system is.

Finally, future research should aim for disentangling the different effects of source and medium when comparing human and computer feedback, ideally comparing all the different combinations possible in one experimental study.

### 4.6. Implications

At the theoretical level, our findings may bear some relevance to the 'Computers are Social Actors' (CASA) paradigm, which suggests that humans consider computers as social actors, interacting with them in the same way as with other humans (Nass et al., 1994; Nass & Moon, 2000; Sundar & Nass, 2000). According to this work, it is sufficient that a computer communicates to a user via text (such as in Study 2) to be considered as a social actor. It is already known that these human-machine interactions can lead to group identification and even peer pressure with computers (Xu & Lombard, 2017). The present article goes beyond that and adds to the CASA paradigm by showing how computers (possibly perceived as social agents) can be just as effective as humans when inducing social stress and influencing emotions. Furthermore, there is something happening at the interpersonal level in these human-machine interactions, as shown in the fairness ratings, which can further highlight how computers are perceived as social agents.

At the practical level, our results highlight some negative consequences of using destructive elements in feedback, regardless of the source from which the feedback stems. Even though performance was unimpaired, stress and negative affect still increased. Therefore, it could be beneficial to train managers or program automated systems to use constructive elements of feedback rather than destructive ones. Automated feedback might require additional carefulness due to how they are perceived. Perceived interpersonal fairness could be taken into account to increase acceptance of such systems in the workplace. For example, if unfairness implies a lack of acceptance of automation, high transparency in how such systems evaluate performance, generate feedback and function in general could possibly help in this direction.

### 4.7. Conclusion

The present studies were one of the first that examined the influence of negative performance feedback on a wide range of established cognitive tests, totaling six different tasks. Study 2 was the first to investigate a modern form of computer negative feedback as a source while controlling for feedback medium. No impairments on subsequent performance were found, regardless of task type and feedback source. This was interpreted as supporting the 'blank out'-mechanism postulated in Sauer et al. (2019), suggesting that participants were able to protect their performance from social stress. Although our results show successful performance maintenance, the findings do not speak in favor of using destructive elements in negative feedback at work, due to the apparent costs of performance maintenance. Indeed, negative feedback caused stress, negative mood and anger. These costs raise some concerns about repeated inductions of social stress and negative feedback in the long term, particularly for computer feedback.

One of the most interesting results of this paper is that negative computer feedback was perceived as more unfair than human feedback. In times when algorithms, artificial intelligence and automation are increasingly prominent not only in the workplace but in society in general, human-machine interaction continues to be a highly relevant topic. In such a context, the finding surrounding fairness, if it were to be replicated in further studies, might have important ramifications for automation design and interactions with humans. Perceived lack of fairness may possibly lead to counterproductive behavior in the workplace, ultimately impairing performance. It would therefore appear wise not to forget employees' perception of an automated system before implementing it in the workplace.

## Credit author statement

Simon Thuillard: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Project administration Mara Adams: Methodology, Investigation, Formal analysis, Writing – original draft Giacomo Jelmini: Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing Sven Schmutz: Conceptualization, Writing – original draft Andreas Sonderegger: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition Juergen Sauer: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Funding

This work was supported by the Swiss National Science Foundation (grant No. 173344).

## Declaration of competing interest

None.

## Acknowledgements

We are very grateful to the Swiss National Science Foundation for funding this work (No. 173344). We would also like to thank Morgane Sottas and Giorgia Trojan for their help in completing this research, and to Marianne Richter for her help at various stages of this research. Finally, we would like to thank three anonymous reviewers for their work and their help in improving the present paper.

## Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.chb.2022.107270.

## References


Alder, G. S. (2007). Examining the relationship between feedback and performance in a monitored environment: A clarification and extension of feedback intervention theory. *The Journal of High Technology Management Research, 17*(2), 157–174. https://doi.org/10.1016/j.hitech.2006.11.004

Alder, G. S., & Ambrose, M. L. (2005). An examination of the effect of computerized performance monitoring feedback on monitoring fairness, performance, and satisfaction. *Organizational Behavior and Human Decision Processes, 97*(2), 161–177. https://doi.org/10.1016/j.obhdp.2005.03.003

Ammons, R. B. (1956). Effects of knowledge of performance: A survey and tentative theoretical formulation. *The Journal of General Psychology, 54*(2), 279–299. https://doi.org/10.1080/00221309.1956.9920284

Ashforth, B. E., & Schinoff, B. S. (2016). Identity under construction: How individuals come to define themselves in organizations. *Annual Review of Organizational Psychology and Organizational Behavior, 3*, 111–137. https://doi.org/10.1146/annurev-orgpsych-041015-062322

Baron, R. A. (1988). Negative effects of destructive criticism: Impact on conflict, self-efficacy, and task performance. *Journal of Applied Psychology, 73*(2), 199–207. https://doi.org/10.1037/0021-9010.73.2.199

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Breyer, B., & Bluemke, M. (2016). Deutsche version der Positive and negative affect Schedule PANAS (GESIS panel). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. https://doi.org/10.6102/zis242

Brickenkamp, R. (1962). *Aufmerksamkeits-belastungs-test (test d2)* (1st ed.). Göttingen, Germany: Hogrefe.

Brown, J. D. (2010). High self-esteem buffers negative feedback: Once more with feeling. *Cognition and Emotion, 24*(8), 1389–1404. https://doi.org/10.1080/02699930903504405

Byrne, K. A., Tibbett, T. P. L., Laserna, N., Carter-Sowell, A. R., & Worthy, D. A. (2016). Ostracism reduces reliance on poor advice from others during decision making. *Journal of Behavioral Decision Making, 29*(4), 409–418. https://doi.org/10.1002/bdm.1886

Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology, 83*(4), 615–633. https://doi.org/10.1037/0021-9010.83.4.615

Cianci, A. M., Klein, H. J., & Seijts, G. H. (2010). The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness. *Journal of Applied Psychology, 95*(4), 618–630. https://doi.org/10.1037/a0019130

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology, 74*(1), 130–135. https://doi.org/10.1037/0021-9010.74.1.130

Colquitt, J. A., Long, D. M., Rodell, J. B., & Halvorsen-Ganepola, M. D. (2015). Adding the "in" to justice: A qualitative and quantitative investigation of the differential effects of justice rule adherence and violation. *Journal of Applied Psychology, 100*(2), 278–294. https://doi.org/10.1037/a0038131

Colquitt, J., Scott, B., Rodell, J., Long, D., Zapata, C., Conlon, D., & Wesson, M. (2013). Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology, 98*, 199–236.

Cortina, J. M., & Folger, R. G. (1998). When is it acceptable to accept a null hypothesis: No way, jose? *Organizational Research Methods, 1*(3). https://doi.org/10.1177/109442819813004

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*, 399–412. https://doi.org/10.1111/bjop.12046

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection tasks. *Human Factors, 44*(1), 79–94. https://doi.org/10.1518/0018720024494856

Earley, P. C. (1988). Computer-generated performance feedback in the magazine-subscription industry. *Organizational Behavior and Human Decision Processes, 41*(1), 50–64. https://doi.org/10.1016/0749-5978(88)90046-5

Eatough, E. M., Meier, L. L., Igic, I., Elfering, A., Spector, P. E., & Semmer, N. K. (2016). You want me to do what? Two daily diary studies of illegitimate tasks and employee well-being. *Journal of Organizational Behavior, 37*, 108–127. https://doi.org/10.1002/job.2032

Fletcher, C. (1997). *Appraisal: Routes to improved performance*. London: Chartered Institute of Personnel & Development.

Fyfe, E. R., & Rittle-Johnson, B. (2016). The benefits of computer-generated feedback for mathematics problem solving. *Journal of Experimental Child Psychology, 147*, 140–151. https://doi.org/10.1016/j.jecp.2016.03.009

Gardner, E. P., Young, P., & Ruth, S. R. (1989). Evolution of attitudes toward computers: A retrospective view. *Behaviour & Information Technology, 8*(2), 89–98. https://doi.org/10.1080/01449298908914542

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology, 60*(6), 895–910. https://doi.org/10.1037/0022-3514.60.6.895

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*, 1280–1300.

Hockey, G. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive–energetical framework. *Biological Psychology, 45*, 73–93.

Hockey, G. R. J., & Hamilton, P. (1983). The cognitive patterning of stress states. In G. R. J. Hockey (Ed.), *Stress and fatigue in human performance* (pp. 331–362). New York: Wiley.

Holbrook, R. (2002). Contact points and flash points: Conceptualizing the use of justice mechanisms in the performance appraisal interview. *Human Resource Management Review, 12*(1), 101–123. https://doi.org/10.1016/S1053-4822(01)00053-5

Immonen, M., & Sintonen, S. (2015). Evolution of technology perceptions over time. *Information Technology & People, 28*(3), 589–606. https://doi.org/10.1108/ITP-12-2013-0219

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'trier social stress test' – a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*, 76–81. https://doi.org/10.1159/000119004

Kluger, A. N., & Adler, S. (1993). Person-versus computer-mediated feedback. *Computers in Human Behavior, 9*(1), 1–16. https://doi.org/10.1016/0747-5632(93)90017-M

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Krings, R., Jacobshagen, N., Elfering, A., & Semmer, N. K. (2015). Subtly offending feedback. *Journal of Applied Social Psychology, 45*, 191–202. https://doi.org/10.1111/jasp.12287

Kuvaas, B. (2006). Performance appraisal satisfaction and employee outcomes: Mediating and moderating roles of work motivation. *International Journal of Human Resource Management, 17*(3), 504–522. https://doi.org/10.1080/09585190500521581

Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *STAI-das state-trait-angstinventar*. Weinheim: Beltz.

Lechermeier, J., & Fassnacht, M. (2018). How do performance feedback characteristics influence recipients' reactions? A state-of-the-art review on feedback source, timing, and valence effects. *Management Review Quarterly, 68*(2), 145–193. https://doi.org/10.1007/s11301-018-0136-8

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In *Paper presented at the proceedings of the 33rd annual ACM conference on human factors in computing systems - CHI '15*.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103.

Lustenberger, D. E., & Jagacinski, C. M. (2010). Exploring the effects of ostracism on performance and intrinsic motivation. *Human Performance, 23*(4), 283–304. https://doi.org/10.1080/08959285.2010.501046

McFarlin, D. B., & Blascovich, J. (1984). On the Remote Associates test (rat) as an alternative to illusory performance feedback: A methodological note. *Basic and Applied Social Psychology, 5*(3), 223–229. https://doi.org/10.1207/s15324834basp0503_5

Mednick, S. A. (1968). The Remote Associates test. *Journal of Creative Behavior, 2*(3), 213–214.

Moore, D. A., & Klein, W. M. P. (2008). Use of absolute and comparative performance feedback in absolute and comparative judgments and decisions. *Organizational Behavior and Human Decision Processes, 107*(1), 60–74. https://doi.org/10.1016/j.obhdp.2008.02.005

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: Sage.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*, 81–103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 72–78). New York, NY: ACM.

Nease, A. J. A., Mudgett, B. O., & Quinones, M. A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of Applied Psychology, 84*(5), 806–814. https://doi.org/10.1037//0021-9010.84.5.806

Nebeker, D. M., & Tatum, B. C. (1993). The effects of computer monitoring, standards, and rewards on work performance, job-satisfaction, and stress. *Journal of Applied Social Psychology, 23*(7), 508–536. https://doi.org/10.1111/j.1559-1816.1993.tb01101.x

Nummenmaa, L., & Niemi, P. (2004). Inducing affective states with success-failure manipulations: A meta-analysis. *Emotion, 4*(2), 207–214. https://doi.org/10.1037/1528-3542.4.2.207

O'Connor, A. J., Nemeth, C. J., & Akutsu, S. (2013). Consequences of beliefs about the malleability of creativity. *Creativity Research Journal, 25*(2), 155–162. https://doi.org/10.1080/10400419.2013.783739

Onnasch, L. (2015). Crossing the boundaries of automation – function allocation and reliability. *International Journal of Human-Computer Studies, 76*, 12–21. https://doi.org/10.1016/j.ijhcs.2014.12.004

Ötting, S. K., & Maier, G. W. (2018). The importance of procedural justice in Human–Machine Interactions: Intelligent systems as new decision agents in organizations. *Computers in Human Behavior, 89*, 27–39. https://doi.org/10.1016/j.chb.2018.07.022

Parasuraman, R., & Miller, C. A. (2004). Trust and etiquette in high-criticality automated systems. *Communications of the ACM, 47*(4), 51–55. https://doi.org/10.1145/975817.975844

Peifer, C., Sauer, J., & Antoni, C. H. (2020). Effects of social stress on performance and strain in complex multiple task environments. *Ergonomics*. https://doi.org/10.1080/00140139.2020.1765028

Promberger, M., & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making, 19*(5), 455–468. https://doi.org/10.1002/bdm.542

Raver, J. L., Jensen, J. M., Lee, J., & O'Reilly, J. (2012). Destructive criticism revisited: Appraisals, task outcomes, and the moderating role of competitiveness. *Applied Psychology-an International Review-Psychologie Appliquee-Revue Internationale, 61*(2), 177–203. https://doi.org/10.1111/j.1464-0597.2011.00462.x

Resnik, P. V., & Lammers, H. B. (1985). The influence of self-esteem on cognitive responses to machine-like versus human-like computer feedback. *The Journal of Social Psychology, 125*(6), 761–769. https://doi.org/10.1080/00224545.1985.9713550

Runco, M. A., Illies, J. J., & Reiter-Palmon, R. (2005). Explicit instructions to Be creative and original: A comparison of strategies and criteria as targets with three types of divergent thinking tests. *Korean Journal of Thinking and Problem Solving, 15*(1), 5–15.

Sauer, J., Schmutz, S., Sonderegger, A., & Messerli, N. (2019). Social stress and performance in human-machine interaction: A neglected research field. *Ergonomics, 62*(11), 1377–1391. https://doi.org/10.1080/00140139.2019.1652353

Sauer, J., Smargiassi, O., Jeanneret, A., & Thuillard, S. (2020). *Human and machine-induced social stress and cognitive performance*. *Ergonomics.* Informa UK Limited. https://doi.org/10.1080/00140139.2020.1850883

Schulte-Braucks, J., Baethge, A., Dormann, C., & Vahle-Hinz, T. (2019). Get even and feel good? Moderating effects of justice sensitivity and counterproductive work behavior on the relationship between illegitimate tasks and self-esteem. *Journal of Occupational Health Psychology, 24*(2), 241–255. https://doi.org/10.1037/ocp0000112

Semmer, N. K., Jacobshagen, N., Meier, L. L., & Elfering, A. (2007). Occupational stress research: The stress-as-offense-to-self perspective. In S. McIntyre, & J. Houdmont (Eds.), *Occupational health psychology: European perspectives on research, education and practice* (Vol. 2, pp. 41–58). Nottingham: Nottingham University Press.

Semmer, N. K., Tschan, F., Jacobshagen, N., Beehr, T. A., Elfering, A., Kälin, W., & Meier, L. L. (2019). Stress as Offense to self: A promising approach comes of age. *Occupational Health Science, 3*(3), 205–238. https://doi.org/10.1007/s41542-019-00041-5

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. V. (1983). *Manual for the state-trait anxiety inventory*. Consulting Psychologists Press.

Stanton, J. M. (2000). Reactions to employee performance monitoring: Framework, review, and research directions. *Human Performance, 13*(1), 85–113. https://doi.org/10.1207/s15327043hup1301_4

Steinborn, M. B., Langner, R., Flehmig, H. C., & Huestegge, L. (2018). Methodology of performance scoring in the d2 sustained-attention test: Cumulative-reliability functions and practical guidelines. *Psychological Assessment, 30*(3), 339–357. https://doi.org/10.1037/pas0000482

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor? *Communication Research, 27*, 683–703.

Thuillard, S. and Richter, M. ((Unpublished results)).

Van Dijk, D., & Kluger, A. N. (2011). Task type as a moderator of positive/negative feedback effects on motivation and performance: A regulatory focus perspective. *Journal of Organizational Behavior, 32*(8), 1084–1105. https://doi.org/10.1002/job.725

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.

Wechsler, D. (1981). *WAIS-R : Wechsler adult intelligence scale-revised*. New York, N.Y.: Psychological Corporation.

Wesche, J. S., & Sonderegger, A. (2019). When computers take the lead: The automation of leadership. *Computers in Human Behavior, 101*, 197–209. https://doi.org/10.1016/j.chb.2019.07.027

Xu, K., & Lombard, M. (2017). Persuasive computing: Feeling peer pressure from multiple computer agents. *Computers in Human Behavior, 74*, 152–162.