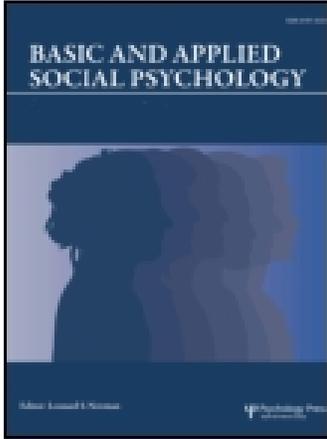


This article was downloaded by: [Universitaire De Lausanne]

On: 14 February 2015, At: 06:04

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Basic and Applied Social Psychology

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/hbas20>

### The Effect of Grades on the Preference Effect: Grading Reduces Consideration of Disconfirming Evidence

Anne-Sophie Hayek<sup>a b</sup>, Claudia Toma<sup>c</sup>, Dominique Oberlé<sup>b</sup> & Fabrizio Butera<sup>a</sup>

<sup>a</sup> Université de Lausanne

<sup>b</sup> Université de Paris Ouest Nanterre La Défense

<sup>c</sup> Tilburg Universiteit

Published online: 04 Nov 2014.



CrossMark

[Click for updates](#)

To cite this article: Anne-Sophie Hayek, Claudia Toma, Dominique Oberlé & Fabrizio Butera (2014) The Effect of Grades on the Preference Effect: Grading Reduces Consideration of Disconfirming Evidence, *Basic and Applied Social Psychology*, 36:6, 544-552, DOI: [10.1080/01973533.2014.969840](https://doi.org/10.1080/01973533.2014.969840)

To link to this article: <http://dx.doi.org/10.1080/01973533.2014.969840>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# The Effect of Grades on the Preference Effect: Grading Reduces Consideration of Disconfirming Evidence

Anne-Sophie Hayek

*Université de Lausanne and Université de Paris Ouest Nanterre La Défense*

Claudia Toma

*Tilburg Universiteit*

Dominique Oberlé

*Université de Paris Ouest Nanterre La Défense*

Fabrizio Butera

*Université de Lausanne*

The tendency to look for evidence that supports, rather than questions, one's viewpoint (preference effect) is a pervasive phenomenon. Although one important goal of education is developing critical thinking, the widespread practice of grading might discourage students in appreciating disconfirming evidence. We hypothesized that individual grading increases the preference effect. In Experiment 1, participants who expected to be graded exhibited a higher preference effect compared to participants who expected their work to be merely visible. Experiment 2 replicated this effect and further showed that grading increased participants' perception of a competitive social comparison. Implications for educational policies are discussed.

In many educational systems, students learn that in order to write a convincing essay, they must include a thesis, an antithesis, and a synthesis in the development of their argument. The rationale for this recommendation is that once students have stated their main point of view, they should be able to refute it, or to propose an alternative point of view, and then to come up with a perspective that includes, compares, and articulates the opposing points of view. This procedure, however, requires from students to be able to decenter from a single idea or hypothesis, that is, to question their own point of view, an ability that is easily impaired when students are under some evaluative pressure (Butera & Buchs, 2005). In an environment where evaluative pressure is pervasive, as it is the case with grading at school and university, it is then possible that students want to

confirm their point of view rather than being open to information that might question their position. In the present research, we aim at addressing this problem and testing the effects of grades on people's tendency to look for evidence that confirms initial preferences. We hypothesized that the expectation of being graded will increase this tendency compared to situations in which people's work is simply made visible.

## GRADES AND PROPERTIES OF GRADES

Grades can be used to produce two types of assessment (Brookhart, 2004). On one hand, they allow comparing the current level of performance (or knowledge) of a person to given criteria (i.e., criterion-referenced assessments). On the other hand, they can be used to compare levels of performance (or knowledge) across individuals, thereby allowing the establishment of an implicit or

---

Correspondence should be sent to Fabrizio Butera, Université de Lausanne - IP-SSP, Géopolis, CH 1015 Lausanne, Switzerland.  
E-mail: fabrizio.butera@unil.ch

explicit ranking (i.e., norm-referenced assessment; Glaser, 1963). In both cases, the main advantage of grades is the visibility they provide: They summarize performance in a number—or a letter, or a judgment—and thereby constitute an easily interpretable criterion of success (or failure). This is probably the reason why grades constitute the main method of assessment in educational and professional settings (Knight & Yorke, 2003).

The visibility afforded by grades, however, may also originate an undesired by-product: By making very clear the differences in merit across people, grades operate a switch in individuals' interest from a focus on the task to a focus on the social comparison of competences (Pulfrey, Buchs, & Butera, 2011), with detrimental consequences for learning and performance. Indeed, the literature on attentional focusing, conflict regulation, and achievement goals shows that grades could induce self-evaluation threat (Muller & Butera, 2007) and create an evaluative pressure that focuses individuals on performance relative to others (Butler, 1987), which in turn moves individual's attention away from the concerns of mastering the task as it fosters assertion of one's competence over that of others (Darnon & Butera, 2007; Darnon, Muller, Schragger, Pannuzzo, & Butera, 2006). In this respect, an impressive amount of research, dating back more than 20 years, has shown that normative assessment entails a long list of nefarious effects for learning and performance. Grades hinder improvement from one test to the following (R. G. Williams, Pollack, & Ferguson, 1975), they reduce interest in the task at hand (Harackiewicz, Abrahams, & Wageman, 1987), they impair intrinsic motivation and performance (Butler, 1987, 1988; Butler & Nisan, 1986; Pulfrey, Darnon, & Butera, 2013). These results have been replicated and extended by many other research teams, in both psychology (Deci & Ryan, 1985; Harter, 1978) and educational sciences (Black & Wiliam, 1998; Thomas & Oldfather, 1997).

## GRADES AND PREFERENCE EFFECT

One might wonder why is it important to understand the effects of grades on the tendency to look for confirmation of one's own point of view. At least two reasons can be mentioned. First, confirmation bias—the tendency to look for evidence that supports, rather than questions, one's hypothesis or viewpoint—is a pervasive phenomenon, long known to hamper people's ability to develop critical thinking and logical argumentation (Klayman & Ha, 1987). In particular, it has been argued that confirmation is mainly used when people need to defend their point of view from an opponent or from the risk of being wrong, which impairs the ability to

consider alternatives (Butera & Mugny, 2001; Mercier & Sperber, 2011). At the same time, some results suggest that standard methods of teaching (e.g., encouraging students to present reasons for opinions they hold rather than reasons against them) and standard methods of evaluation (e.g., using grades) may foster this bias (Nickerson, 1998). For example, when being graded for written essays, students use more claims that contain supporting evidence than claims that contain disconfirming evidence (Narveson, 1980).

Second, confirmation bias is frequent in groups, which are often used in educational settings. When occurring in groups, this bias (also called preference effect) refers to insufficient revisions of individual preference during group discussions (Brodbeck, Kerschreiter, Mojzisch, Frey, & Schulz-Hardt, 2002). More specifically, the preference effect occurs because group members have the tendency to evaluate information that is consistent with their initial preferences more favorably than information that is inconsistent (Greitemeyer & Schultz-Hardt, 2003). In the domain of group decision making, research has shown that the preference effect increases in situations where individuals try to prove themselves in front of others, for example, in competition (Toma & Butera, 2009; Toma, Gilles, & Butera, 2013). In their recent research, Toma, Gilles, et al. (2013) manipulated members' goals using either an individual (negative goal interdependence—i.e., competition) or a group promotion (positive goal interdependence—i.e., cooperation) that was offered to group members who succeeded in solving the mystery of a car accident case. Participants were also told that other (fictitious) members had either dissenting or identical initial preferences to their own. Results indicated that the preference effect was higher in competition than in cooperation, and especially when participants were facing the dissenting preferences of the other group members. This effect was mediated by self-enhancement strategies, which are known to reflect strivings to raise one's positive self-view and superiority (Alicke & Sedikides, 2009).

The aim of the present research is to determine whether a similar effect could be obtained with grades. At school, students are often defending different points of view and competing with one another for better grades, even when working on group projects. Such practices, however, can have opposite effects. On one hand, this could be motivating because it increases students' visibility (Cameron & Pierce, 2002) and signals a situation in which grades are used to produce criterion-referenced evaluation (i.e., evaluation of a student in comparison with a certain level of knowledge or standard). On the other hand, this could be threatening because it increases comparability of one's work with that of others (Marshall & Weinstein, 1984) and signals a situation in which grades are used to produce

norm-referenced evaluations (i.e., evaluation of a student in comparison with other students). Studies have shown that contexts in which one needs to prove oneself in front of others lead to self-evaluative threats (Dickerson, Gruenewald, & Kemeny, 2004), which in turn increases the preference effect (Toma, Bry, & Butera, 2013; Toma, Gilles, et al., 2013). This should not be the case when visibility is merely emphasized by the presence of a third person (Dickerson, Mycek, & Zaldivar, 2008) or when the self-evaluation threat linked to the normative facet of grades disappears, for example, because one is assured of one's own superiority (Muller & Butera, 2007).

In sum, the visibility afforded by grades may not be a problem in itself. It is rather the potentially competitive social comparison elicited by grades that may focus individuals on the defence of their own point of view. Therefore in the present research we test the hypothesis that in a group situation in which members face the dissent of others, expecting one's work to be graded should increase the preference effect compared to expecting one's work to be merely visible.

## OVERVIEW OF THE STUDIES

Two studies were conducted to test this hypothesis. The task used was the same as the one used by Toma, Gilles, et al. (2013). In this task participants were asked to individually solve a car accident case and to find the person responsible for it (initial preference). Then they were informed that other team members supported different initial preferences because of the different information they possessed. Participants were asked to read and to rate the importance of the other members' information, which was either consistent or inconsistent with their own initial preference. The preference effect was calculated as the difference between the evaluation of consistent and inconsistent information.

In Experiment 1, we tested our main hypothesis that individual grading during group work should increase the preference effect as compared with mere visibility of one's work. To test this hypothesis, we needed to compare the focal condition of evaluation by grades with two control conditions. First, we introduced a control condition in which individual work was simply visible but not graded. Second, we also introduced a control condition without grades or visibility. Indeed, as previously noted, grades imply both a focus on competitive social comparison and increased visibility. If, as hypothesized, grades increase the preference effect because of the social comparison component, the condition with individual grades should differ from the condition with mere visibility. It is, however, possible that, contrary to our expectation, it is indeed visibility

that increases the preference effect; in this case both the graded and the mere visibility conditions should induce a higher preference effect than the control condition without grades or visibility. In Experiment 2, we aimed at replicating Experiment 1 in a more naturalistic context (i.e., classroom) and disentangling possible confounds in the manipulation of visibility.

## EXPERIMENT 1

### Method

#### *Participants*

A total of 61 university students from a Swiss university took part in this experiment. Six participants were excluded from the analyses because they did not comply with the experimental script (i.e., they did not choose Mr. X as the initial preference, to whom the script and clues oriented them). The remaining sample included 55 university students (34 women and 20 men, one student did not mention her or his gender and age;  $M = 18.09$  years,  $SD = 1.20$ ). Participants were randomly assigned to one of three experimental conditions: graded-visible ( $n = 13$ ), nongraded-visible ( $n = 23$ ) and nongraded-nonvisible ( $n = 19$ ). Preliminary analyses revealed that gender did not influence our effects, and therefore this variable was not included in final analyses.

#### *Procedure*

Participants were told that they would participate in a study on the resolution of criminal investigations (for the exact materials, see Toma, Gilles, et al., 2013). They worked individually, and then they were led to imagine that they would work in a team with two other students. The role-play story had participants pretend they worked as police inspectors with two other people in order to identify the party responsible for a car accident. Four people were potential suspects, but three of them could be exonerated (Mr. X, Mrs. Y, Mr. Z) and the fourth (Mr. X's son) incriminated based on a critical set of nine clues. All participants were oriented toward Mr. X, whereas the two other fictitious members were attributed Mrs. Y and Mr. Z as initial preferences. All participants were asked to commit to their initial preference (Mr. X), and then they were informed that they did not possess the entire set of information and that for this reason they would be provided with supplementary information given by the two other participants.

The manipulation of grades was introduced at this point. In the graded-visible condition, participants were told that the chief would be present and give an individual grade (ranging from 1 to 6, which corresponds to the

usual grading range in Switzerland) to each inspector at the end of the investigation. In the nongraded-visible condition, participants were told that the chief would be present because he is interested to follow the investigation. In the nongraded-nonvisible condition, participants were told that the chief of the police inspectors would not be able to follow the investigation. It should be noted that the script did not specify that grades would be used as norm-based assessment, but we assumed that the participants would behave as a function of this meaning of grades to the extent that norm-based assessment is by far the most used (Knight & Yorke, 2003).

Subsequently, participants received six items of information, three consistent and three inconsistent with their initial preference (Mr. X), supposedly coming from the two other people. An example of consistent information was that the person responsible for the accident is a man. An example of inconsistent information was that the person responsible for the accident is younger than 30 years old (participants knew that Mr. X is 53 years old). The consistent and inconsistent information was presented in random order for each participant. Participants were asked to evaluate the items of information with regard to their importance in making an optimal decision. Finally, participants were asked to make a final decision based on all the pieces of information. At the end, participants were debriefed and thanked for their participation.

### Measures

**Manipulation check.** To check whether participants correctly perceived the graded-visible condition compared to the other conditions, they were asked to answer the following question: *During the investigation, did the chief tell you that you would be individually evaluated?* (Yes/No).

**Preference for consistent information.** Participants evaluated to what extent the six items of information they received were important in reaching the optimal decision, on a scale ranging from 1 (*not at all important*) to 9 (*very important*). This information included three items consistent and three items inconsistent with the participant's initial preference. One consistent item and one inconsistent item were dropped from the analyses, as they lowered the reliability test.<sup>1</sup> The preference for consistent information was computed by subtracting the ratings of the two inconsistent information items

<sup>1</sup>The reliability of the three inconsistent items was  $\alpha = .34$ ; after dropping the concerned item, the reliability raised to  $\alpha = .63$ . The reliability of the three consistent items was  $\alpha = .27$ ; after dropping the concerned item, the reliability raised to  $\alpha = .71$ .

( $r = .47$ ) from the ratings of the two consistent information items ( $r = .55$ ) and refers to the extent to which participants evaluated information in a way that confirmed their initial preference. A positive score indicates that consistent information was considered more valuable than inconsistent information, and a negative score indicates that inconsistent information was considered more valuable than consistent information.

## Results

### Manipulation Check

All participants in the graded-visible condition responded "Yes" and all participants in the other two conditions responded "No" to the question regarding the individual evaluation.

### Preference for Consistent Information

The results suggested that the preference for consistent information was higher in the graded-visible condition ( $M = 1.46$ ,  $SD = 2.00$ ) compared to the nongraded-visible condition ( $M = -0.24$ ,  $SD = 2.09$ ,  $d = 0.65$ ), as predicted, but also slightly higher compared to the nongraded-nonvisible condition ( $M = 0.86$ ,  $SD = 2.13$ ,  $d = 0.29$ ). Of interest, the preference for consistent information was higher in the nongraded-nonvisible condition compared to the nongraded-visible condition ( $d = 0.47$ ). The results are presented in Figure 1. It should be noted that in the whole article we do not report the significance tests, in accordance with *Basic and Applied Social Psychology* policies (Trafimow, 2014; see also Trafimow, 2003).

Although our main interest was to study preference for consistent information, it is common practice in this literature to also report confirmatory decisions. Therefore, a dichotomous measure was derived from the final

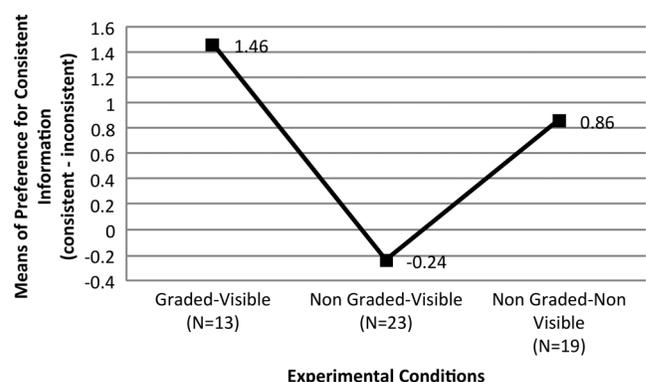


FIGURE 1 Experiment 1: Mean preference for consistent information as a function of the experimental conditions.

decision reported by participants. When the answer confirmed their initial preference (Mr. X), it was coded 1, whereas when the answer did not (Mrs. Y, Mr. Z, or Mr. X's son) it was coded 0. Across all conditions, 49.1% of participants confirmed their initial preference (Mr. X). In particular, the proportion of confirmatory decisions was 10.9% in the graded-visible condition, 25.5% in the nongraded-visible condition, and 12.7% in the nongraded-nonvisible condition.

## Discussion

The results of this experiment provide preliminary evidence for our focal hypothesis. Indeed, participants who expected their work to be individually graded exhibited a higher preference for consistent information compared to participants who expected their work to be only visible. The alternative hypothesis that the preference effect should increase as a function of visibility was not supported.

An interesting, although unexpected, result was that the preference effect was higher in the nongraded-nonvisible condition compared to the nongraded-visible condition. This may suggest that when one's work is not expected to be visible, individuals are not particularly motivated to revise their preferences. This is consistent with research on social loafing showing that people reduce their individual contribution to group work when their effort is not visible (e.g., Karau & Williams, 1993; Latané, Williams, & Harkins, 1979; Williams, Harkins, & Latané, 1981).

## EXPERIMENT 2

The first aim of Experiment 2 was to replicate the effect found in Experiment 1 and to confirm that individual grading indeed increases the preference effect. The second aim was to test an assumption underlying our general hypothesis. We argued in the theoretical introduction that one important component of grading is the focus on competitive comparison with other group members; we therefore added a measure of perceived competition to test whether participants perceived more competition in the graded-visible condition compared to other control conditions (see next).

The third aim was to disentangle possible confounds related to the manipulation of visibility. In the nongraded-visible condition of Experiment 1 participants were told that their chief would be present because he was interested to follow the investigation. It is therefore difficult to know whether the reduced extent of the preference for consistent information was due to the expectation of one's work being visible or to the mere presence of the chief. Thus, in this second experiment

we broke down the former nongraded-visible condition into a condition of visibility and a condition of mere presence. Visibility, or social visibility, is at stake whenever an individual is observed while achieving a task (Bond & Titus, 1983; Zajonc, 1965), whereas mere presence of a person occurs when this person is physically present during the individual's performance and the individual knows that this person is not interested in watching performance (Cottrell, Wack, Sekerak, & Rittle, 1968). In both cases, some extent of visibility of the individual who is performing is at stake, but they are conceptually different. Of interest, such situations that increase individual visibility are only deleterious to the extent that the observer's presence explicitly implies the possibility of a negative evaluation (e.g., when a panel of evaluators is there to observe in a critical and rejecting manner the individual perform; Gruenewald, Kemeny, Aziz, & Fahey, 2004). Otherwise, visibility situations from which explicit negative social evaluation is absent do not elicit stress (Dickerson et al., 2008). Thus, although in both cases of visibility and mere presence individual visibility is at stake, we should not expect any deleterious effect, that is any increase of the preference effect: Both the visibility and the mere presence conditions should then lead to a lower preference effect compared to the graded-visible condition.

Another confound in the nongraded-visible condition was due to the presence of accountability demands. Accountability, defined as the "pressure to justify one's causal interpretation to others" (Tetlock, 1985, p. 227), is often manipulated by increasing the visibility of one's judgement or position taken. For example, participants may be told that they will have to justify their judgement or position to someone else (Tetlock, 1985), which implies both visibility and evaluation. However, contrary to a situation of grading, this evaluation is not normative: The focus of accountability is not on the comparability of one's work to that of others but on the underlying reasons that justify one's own position, on examining and evaluating all available information that would help finding the most appropriate solution (Johnson & Johnson, 1985). Indeed, the literature investigating accountability has shown that asking people to be accountable for their judgments and decisions has positive effects on several tasks, specifically by reducing reasoning biases (Lerner, Goldberg, & Tetlock, 1998; Tetlock, 1983, 1985) and results, for instance, in individuals producing more integrative complex thoughts (Green, Visser, & Tetlock, 2000; Tetlock, 1983), or in becoming more responsive to additional diagnostic evidence (Tetlock & Boettger, 1989). Thus, compared to a grading situation, we expected accountability to reduce the preference effect.

In sum, we hypothesized that the preference effect will be higher in the graded-visible condition compared

to the other three experimental conditions (visibility, mere presence, accountability).

## Method

### Participants

A total of 61 university students from a Swiss university took part to this experiment (42 women and 19 men;  $M = 21.31$  years,  $SD = 1.84$ ). Participants were randomly assigned to one of the four experimental conditions: graded-visible ( $n = 18$ ), visibility ( $n = 16$ ), mere presence ( $n = 13$ ), accountability ( $n = 14$ ). Again gender had no effects and was not included in the final analyses.

### Procedure

The same procedure as in Experiment 1 was used. However, this time the experiment was not conducted in the lab but was part of a class exercise. The instructions used in the experimental conditions were the following: in the graded-visible condition, participants were told that the chief would be present and give each inspector an individual grade (ranging from 1 to 6) at the end of the investigation. In the visibility condition, participants were told that the chief would be present because he is interested to follow the investigation. In the mere presence condition, participants were only told that the chief of the police inspectors would be present. In the accountability condition, participants were told that the chief would be present because the inspectors will have to justify their final decision to him.

Again, participants evaluated the importance of consistent and inconsistent information for the final decision. Finally, they assessed their perception of competition and were debriefed and thanked for their participation.

### Measures

**Preference for consistent information.** This measure was computed as in Experiment 1.<sup>2</sup> The correlation between the two items of consistent information was  $r = .39$ , and the correlation between the two items of inconsistent information was  $r = .58$ .

**Perceived competition.** To test whether indeed the graded-visible condition induced perceived competition to a higher extent than the other experimental conditions, participants answered a two-item questionnaire on a scale ranging from 1 (*not at all*) to 9 (*totally*). The

<sup>2</sup>The reliability of the three inconsistent items was  $\alpha = .46$ ; after dropping the concerned item, the reliability raised to  $\alpha = .73$ . The reliability of the three consistent items was  $\alpha = .19$ ; after dropping the concerned item, the reliability raised to  $\alpha = .56$ .

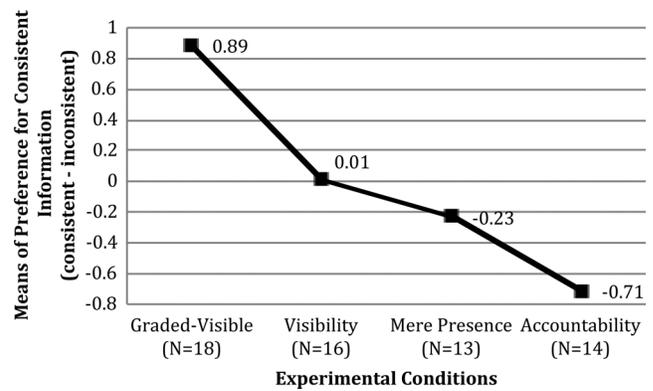


FIGURE 2 Experiment 2: Mean preference for consistent information as a function of the experimental conditions.

questions asked whether participants perceived a competitive atmosphere and whether they perceived the other members as rivals. The two questions were combined into a single score of perceived competition ( $r = .56$ ;  $M = 3.89$ ,  $SD = 1.94$ ).

## Results

### Preference for Consistent Information

The results suggested that the preference for consistent information was higher in the graded-visible condition ( $M = 0.89$ ,  $SD = 1.72$ ) compared to the visibility condition ( $M = 0.01$ ,  $SD = 2.22$ ,  $d = 0.44$ ), the mere presence condition ( $M = -0.23$ ,  $SD = 1.82$ ,  $d = 0.63$ ), and the accountability condition ( $M = -0.71$ ,  $SD = 2.14$ ,  $d = 0.82$ ). The results are presented in Figure 2.

Again, we reported confirmatory decisions. Across all conditions, 31.1% confirmed their initial preference (Mr. X); in particular, the proportion of confirmatory decisions was 9.8% in the graded-visible condition, 9.8% in the visibility condition, 3.3% in the mere presence condition, and 8.2% in the accountability condition.

### Perceived Competition

The results on the perception of competition mirrored the results on the preference for consistent information: Perception of competition was higher in the graded-visible condition ( $M = 4.75$ ,  $SD = 1.96$ ) compared to the visibility condition ( $M = 3.81$ ,  $SD = 2.02$ ,  $d = 0.47$ ), the mere presence condition ( $M = 3.19$ ,  $SD = 1.8$ ,  $d = 0.98$ ), and the accountability condition ( $M = 3.54$ ,  $SD = 1.75$ ,  $d = 0.65$ ).

## Discussion

The first aim of this second experiment was fulfilled, as the results replicated the effect found in Experiment 1.

Indeed, the graded-visible condition increased the preference for consistent information, compared to the other conditions (visibility, mere presence, accountability). Of interest, the effects on perceived competition mirrored the effects on the preference for consistent information, suggesting that participants were more inclined to perceive competition with others when evaluated with grades than in the other experimental conditions. This supports our assumption that evaluation by grades contains a normative component to a larger extent than the other conditions.

## GENERAL DISCUSSION

In many educational systems, the goal of developing critical thinking and logical argumentation among students is often hampered by the use of standard methods of teaching and evaluation, in particular normative assessment under the form of grades. The practice of using grades was shown to hinder students' ability to develop counter-argumentation and to avoid confirmatory tendencies (Nickerson, 1998). Two experiments tested the hypothesis that the expectation of being graded in a group-work situation increases the preference for information that is consistent with one's initial solution.

Experiment 1 showed that participants who expected their work to be individually graded exhibited a higher preference effect compared to participants who expected their work to be merely visible in the group. Experiment 2 replicated this effect and showed that when participants expect their work to be visible, the presence of grades was associated with an increase in the preference effect as compared with the mere visibility of one's work, the mere presence of an evaluative agent, or the expectancy of being accountable for one's work. This second experiment also showed that participants perceived more competition with other group members when they expected to be graded than when in the three other conditions.

Taken together, the two experiments point to the fact that the potential of grades to elicit a preference effect is not due to their visibility component, as suggested by the difference between the graded condition and other visibility conditions in both experiments. A possible interpretation is that this effect is due to the ability of grades to induce a competitive social comparison, as suggested by the result found in Experiment 2 where grades increased the perception of a competitive atmosphere and of others as competitors. To support such an interpretation, future research should directly manipulate what we assume to be at the core of the facilitating effects of grades on the preference effect, namely, a threatening social comparison. In this respect, it is important to note

that social comparison is a pervasive phenomenon, occurring even unconsciously (Mussweiler, Rüter, & Epstude, 2004), and we think grades can produce the aforementioned effects even when institutional grading practices do not provide official opportunities for social comparison (e.g., posting grades or handing out tests in front of the class). Indeed, students can directly or indirectly compare grades during course chitchatting, breaks, and recess but can also be requested to construe relative positions within the class during interaction with parents or other significant adults.

It could be objected that the task we used is quite an unusual one for students, especially with a grade attached to it, and therefore the present results would have low ecological validity. It is indeed true that our specific materials are different from the usual pedagogical materials that can be found at school or university. However, it is important to note that the hidden profile task was created and conceptualized to precisely represent the common situation of work in heterogeneous groups, like students who have different resources or different domains of expertise and are asked to cooperate on a group project (Greitmeyer & Schulz-Hardt, 2003; Stasser & Titus, 1985; Toma, Gilles, et al., 2013). Moreover, the present laboratory studies are arguably less involving than real-life situations in which grades have concrete and long-term consequences on pupils' and students' prospects in their curriculum and eventually in their professional life. Hence, given the majority of medium to large effect sizes observed on rather small samples (most *ds* ranged from 0.44 to 0.82), we would expect that with usual scholastic materials the results observed in this study should be replicated and the effect sizes should be even larger.

Finally, the present research has disentangled the effects of grades from the effects of visibility, and in this respect it has showed that the effects of grades on the preference effect are most probably due to their normative facet, that is, the fact that they facilitate competitive social comparison with coworkers and not to their visibility facet. This general result has the potential to stimulate new research in two directions. First, the present results point to the fact that the use of grades, possibly due to the focus on competition with other group members, interferes with the capacity of individuals to consider, in an unbiased way, information coming from others. This is a highly valued capacity, especially in decision-making contexts where important decisions need to be made and where, precisely, valuing different alternatives can help fighting pressure to conformity and avoiding situations of defective group decision making in which alternative options are often not considered or too rapidly rejected, that is, the well-known groupthink phenomenon (Janis, 1982). Hence, the present results show that grades can increase

such biased individual appreciation of new information presentation, which could intervene in more complex group decision-making situations, including cooperative ones. It also follows from this discussion that grades may nullify the dynamics of information exchange, a skill that is extremely important in educational practices involving group learning (Johnson & Johnson, 1985).

Second, these results may also inform the scientific community as regards its own functioning, and in particular its capacity to put scientific knowledge into question. Indeed, notwithstanding Popper's (1966) call for considering that scientific theories should be falsifiable, scientists have long been shown to indulge in confirmation bias just as lay people (Mahoney, 1976; Mitroff, 1974). The present study paves the way for future applied studies that could aim at investigating why scientists are often reluctant to admit that they might be wrong. One possibility, building on the reported results, is that the pressure of academic evaluation—under various forms of normative assessment such as impact factor of one's publications, H-factor of individual researchers, or ranking of one's university—might accentuate the phenomenon of preference for consistent information.

#### ACKNOWLEDGMENTS

This article was part of Anne-Sophie Hayek's doctoral thesis under the supervision of Fabrizio Butera and Dominique Oberlé. We thank Benoît Dompnier for invaluable comments and methodological support; François Bavaud and Dominique Bourg for access to their students; and Benoît Dompnier, Thomas Tüscher, Jung Yin Fang, Emmanuelle Anex, Sofia Guidotti, Maria Mirabela Canciu, and Nathalie Faessler for support in data collection.

#### FUNDING

This work was supported by the Swiss National Science Foundation.

#### REFERENCES

- Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*, 1–48.
- Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London, UK: Granada Learning.
- Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin*, *94*, 265–292.
- Brodbeck, F. C., Kerschreiter, R., Mojzisch, A., Frey, D., & Schulz-Hardt, S. (2002). The dissemination of critical, unshared information in decision-making groups: The effects of pre-discussion dissent. *European Journal of Social Psychology*, *32*, 35–56.
- Brookhart, S. M. (2004). Assessment theory for college classrooms. *New Directions for Teaching and Learning*, *100*, 5–14.
- Butera, F., & Buchs, C. (2005). Reasoning together: From focusing to decentering. In V. Girotto & P. N. Johnson-Laird (Eds.), *The shape of reason* (pp. 193–203). Hove, UK: Psychology Press.
- Butera, F., & Mugny, G. (2001). Conflicts and social influences in hypothesis testing. In C. De Dreu & N. De Vries (Eds.), *Group consensus and minority influence: Implications for innovation* (pp. 160–182). Oxford, UK: Blackwell.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, *79*, 474–482.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*, 1–14.
- Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, *78*, 210–216.
- Cameron, J., & Pierce, W. D. (2002). *Rewards and intrinsic motivation: Resolving the controversy*. Westport, CT: Bergin & Garvey.
- Cottrell, N. B., Wack, D. L., Sekerak, G. Y., & Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology*, *9*, 245–250.
- Darnon, C., & Butera, F. (2007). Learning or succeeding? Conflict regulation with mastery or performance goals. *Swiss Journal of Psychology*, *66*, 145–152.
- Darnon, C., Muller, D., Schrage, S. H., Pannuzzo, N., & Butera, F. (2006). Mastery and performance goals predict epistemic and relational conflict regulation. *Journal of Educational Psychology*, *98*, 766–776.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum.
- Dickerson, S. S., Gruenewald, T. L., & Kemeny, M. E. (2004). When the social self is threatened: Shame, physiology, and health. *Journal of Personality*, *72*, 1191–1216.
- Dickerson, S. S., Mycek, P. J., & Zaldivar, F. (2008). Negative social evaluation, but not mere social presence, elicits cortisol responses to a laboratory stressor task. *Health Psychology*, *27*, 116–121.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, *18*, 519–521.
- Green, M. C., Visser, P. S., & Tetlock, P. E. (2000). Coping with accountability cross-pressures: Low-effort evasive tactics and high-effort quests for complex compromises. *Personality and Social Psychology Bulletin*, *26*, 1380–1391.
- Greitemeyer, T., & Schulz-Hardt, S. (2003). Preference-consistent evaluation of information in the hidden profile paradigm: Beyond group-level explanations for the dominance of shared information in group decisions. *Journal of Personality and Social Psychology*, *84*, 322–339.
- Gruenewald, T. L., Kemeny, M. E., Aziz, N., & Fahey, J. L. (2004). Acute threat to the social self: Shame, social self-esteem, and cortisol activity. *Psychosomatic Medicine*, *66*, 915–924.
- Harackiewicz, J. M., Abrahams, S., & Wageman, R. (1987). Performance evaluation and intrinsic motivation: The effects of evaluative focus, rewards, and achievement orientation. *Journal of Personality and Social Psychology*, *53*, 1015–1023.
- Harter, S. (1978). Pleasure derived from challenge and the effects of receiving grades on children's difficulty level choices. *Child Development*, *49*, 788–799.

- Janis, I. L. (1982). *Groupthink* (2nd ed.). Boston, MA: Houghton-Mifflin.
- Johnson, D. W., & Johnson, R. (1985). Classroom conflict: Controversy versus debate in learning groups. *American Educational Research Journal*, 22, 237–256.
- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65, 681–706.
- Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, 94, 211–228.
- Knight, P., & Yorke, M. (2003). *Assessment, learning and employability*. Maidenhead, UK: Open University Press.
- Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37, 822–832.
- Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin*, 24, 563–574.
- Mahoney, M. J. (1976). *Scientist as subject: The psychological imperative*. Cambridge, MA: Ballinger.
- Marshall, H. H., & Weinstein, R. S. (1984). Classroom factors affecting students' self-evaluations: An interactional model. *Review of Educational Research*, 54, 301–325.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34, 57–74.
- Mitroff, I. (1974). Norms and counter-norms in a select group of Apollo moon scientists: A case study of the ambivalence of scientists. *American Sociological Review*, 39, 579–595.
- Muller, D., & Butera, F. (2007). The focusing effect of self-evaluation threat in coaction and social comparison. *Journal of Personality and Social Psychology*, 93, 194–211.
- Mussweiler, T., Rüter, K., & Epstude, K. (2004). The man who wasn't there: Subliminal social comparison standards influence self-evaluation. *Journal of Experimental Social Psychology*, 40, 689–696.
- Narveson, R. D. (1980). Development and learning: Complementary or conflicting aims in humanities education? In R. G. Fuller et al. (Eds.), *Piagetian programs in higher education* (pp. 79–88). Lincoln, NE: ADAPT Program.
- Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220.
- Popper, K. R. (1966). *Conjectures and refutations*. New York, NY: Basic Books.
- Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology*, 103, 683–700.
- Pulfrey, C., Darnon, C., & Butera, F. (2013). Autonomy and task performance: Explaining the impact of grades on intrinsic motivation. *Journal of Educational Psychology*, 105, 39–57.
- Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48, 1467–1478.
- Tetlock, P. E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly*, 46, 285–292.
- Tetlock, P. E. (1985). A social check on the fundamental attribution error. *Social Psychology Quarterly*, 48, 227–236.
- Tetlock, P. E., & Boettger, R. (1989). Accountability: A social magnifier of the dilution effect. *Journal of Personality and Social Psychology*, 57, 388–398.
- Thomas, S., & Oldfather, P. (1997). Intrinsic motivations, literacy, and assessment practices: That's my grade. That's me. *Educational Psychologist*, 32, 107–123.
- Toma, C., Bry, C., & Butera, F. (2013). Because I'm worth it! (More than others...): Cooperation, competition and ownership bias in group-decision making. *Social Psychology*, 44, 248–255.
- Toma, C., & Butera, F. (2009). Hidden profiles and concealed information: Strategic information sharing and use in group decision making. *Personality and Social Psychology Bulletin*, 35, 793–806.
- Toma, C., Gilles, C., & Butera, F. (2013). Strategic use of preference confirmation in group decision making: The role of competition and dissent. *British Journal of Social Psychology*, 52, 44–63.
- Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's theorem. *Psychological Review*, 110, 526–535.
- Trafimow, D. (2014). Editorial. *Basic and Applied Social Psychology*, 36, 1–2.
- Williams, K., Harkins, S., & Latané, B. (1981). Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology*, 40, 303–311.
- Williams, R. G., Pollack, M. J., & Ferguson, N. A. (1975). Differential effects of two grading systems on student performance. *Journal of Educational Psychology*, 67, 253–257.
- Zajonc, R. B. (1965). Social facilitation. *Science*, 149, 269–274.