Between-Subject or Within-Subject Measures of Regret: Dilemma and Solution

Jiehai Zhang
Shanghai Academy of Social Sciences

Clare Walsh
Brown University

Jean-François Bonnefon
University of Toulouse

The "action effect", according to which actions produce more regret than failures to act, has been shown to disappear in between-subject designs. This phenomenon is replicated in a first study. It is then argued that this disappearance is due to the inability of regret scales to capture differences in perceived regret when used in between-subject designs, a difficulty that is highlighted by a second study. A new method, the *common reference method*, is proposed in order to overcome this problem. This method is demonstrated in a third and fourth study, and its boundary conditions are discussed, together with its possible extensions.

Scales are a method of choice for collecting psychological data. Scales allow more powerful analysis than categorical ratings, for they hold much more information, in terms of order and relative distance between values. Yet the use of scales is not always without problem - in particular, they may easily lead to artefactual results, as we will argue in this article in the context of the psychology of regret.

We begin with a conundrum. Whereas many studies using within-subject comparisons have established that actions are regretted more than inactions, this difference seems to disappear when using between-subject comparisons. We replicate this finding in Experiment 1 and consider the possibility that the use of scales in a between-subject design is methodologically inappropriate to investigate differences in regret.

We suggest that regret scales lack the well-defined reference point which would be needed to make relevant between-subject comparisons. We demonstrate (Experiment 2) that regret scales in a between-subject design cannot capture the uncontroversial difference in regret after a missed opportunity of $1,000 and a missed opportunity of $2,000. We then introduce our common reference method, which is meant to improve the reliability of scale-based, between-subject comparisons. We demonstrate (Experiment 3) that the common reference method indeed leads to a recovery of the action effect in a between-subject design. Experiment 4 rules out an alternative explanation of Experiment 3's results, and shows that the common reference method also leads to the recovery, in a between-subject design, of the difference in regret over Experiment 2's $2,000 miss and $1,000 miss. We finally discuss the implications of our results for the design and interpretation of psychological studies of subjective quantities such as regret.

It will be useful to keep in mind through the article that we wish to make a methodological point about the study of regret, but not a theoretical point about regret itself. As a consequence, we will not provide a full-fledged review of the literature on regret, nor will we be concerned over factors that have been shown to have independent influence on our chief example, the action-inaction effect (e.g., whether the decision is justified by previous experience, Inman & Zeelenberg, 2002, or whether the decision is consistent with the decision maker's risk orientation, Seta, McElroy, & Seta, 2001).

The Action Effect Conundrum

Kahneman and Tversky (1982) used an oft-cited stock story to show that actions lead to more regret than inactions with the same bad outcome. They presented their participants with the following story:

> Paul owns shares in company A. During the past year he considered switching to stock in company B, but he decided against it. He now finds out that he would have been better off by $1,200 if he had switched to the stock of company B. George owned shares in company B. During the past year he switched to stock in company A. He now finds out that he would have been better off by $1,200 if he had kept his stock in Company B. Who feels more regret? (p.173)

Participants judged that George (the actor) would experience more regret than Paul (the non-actor). This phenomenon, which has been referred to as the action effect, has been proven robust through a great number of studies (Byrne &

Send correspondence regarding this article to Zhang Jiehai, Institute of Sociology, Shanghai Academy of Social Sciences, 622-7, Huaihai Zhonglu, Shanghai, 200020, China. E-mail: jiehaiz@sass.org.cn. This research benefited from a grant from the French Ministry of Foreign Affairs. We also express our gratitude to Ciping Deng for his help with data collection, to Stéphane Vautier for fruitful discussions on measurement issues, and to Klaus Fiedler for useful suggestions on improving our initial manuscript.

McEleney, 2000; Connolly, Ordonez, & Coughlan, 1997; Gilovich, Medvec, & Chen, 1995; Gleicher et al., 1990; Landman, 1987; Zeelenberg, van den Bos, van Dijk, & Pieters, 2002).

The first to challenge the finding were N'gbala and Branscombe (1997). They argued that the reason why researchers succeeded in obtaining the action effect was the *constant use of a within-subject design*. In contrast, they suggested that the action effect may disappear in the absence of a direct comparison between the actor and non-actor, that is, in a between-subject design. In their first experiment, N'gbala and Branscombe used Kahneman and Tversky's (1982) stock story both in a within-subject design (participants read about the decisions of both characters, and had to rate the regret felt by those characters on two separate 11-point scales) and in a between-subject design (two groups of participants read either about the decision of the actor or about the decision of the non-actor, and had to rate the regret felt by the character on an 11-point scale).

The mean regret attributed to the actor and to the non-actor, in within-subject design, was 83.0 and 70.83 respectively (a reliable difference). In contrast, in between-subject design, regret ratings for the actor and non-actor were comparable (69.00 vs. 68.00). It thus appears that the action effect indeed disappears when using a between-subject design. N'gbala and Branscombe (1997) concluded that the effect could only occur from a direct comparison between the action and the inaction.

Our aim in this article is to investigate this claim, and ultimately to prove it wrong. We will show that the disappearance of the action effect in a between-subject design is a methodological problem, to which we will provide a methodological solution. However, we begin this article with a tentative replication of this phenomenon.

Experiment 1

Method

Participants. A total of 177 undergraduate students, all native Chinese speakers, took part in the experiment (26 men and 151 women; mean age = 20.3, *SD* = 0.9).

Materials and design. Participants in the *Both Agents* condition were given the usual stock story:

> Paul owns shares in company A. During the past year he considered switching to stock in company B, but he decided against it. He now finds out that he would have been better off by $1,200 if he had switched to the stock of company B. George owned shares in company B. During the past year he switched to stock in company A. He now finds out that he would have been better off by $1,200 if he had kept his stock in Company B.

The order the characters appeared in the story was counterbalanced across questionnaires. Participants had to rate the regret felt by Paul and George, on two separate 11-point scales, ranging from 0 ("no regret at all") to 10 ("much regret").

Table 1
Experiment 1: Means and Standard Deviations in Perceived Regret for the Actor and the Non-actor

	Actor	Non-actor
Both Agents	6.9 (2.8)	4.6 (2.5)
Actor Only	5.1 (2.6)	n/a
Non-actor Only	n/a	4.2 (2.6)

Participants in the *Non-actor Only* and the *Actor Only* conditions were only provided with the information about Paul and George respectively, and had to rate the regret the character felt about his decision, again on an 11-point scale.

Procedure. Questionnaires were administered during a class. The experiment was conducted in Chinese.

Results

Table 1 presents the mean perceived regret for the actor (George) and the non-actor (Paul) in the three experimental groups. As expected, in the *Both Agents* condition, more regret is attributed to the actor ($m = 6.9$) than to the non-actor ($m = 4.6$), $t = 7.01$, $p < .001$ (one-tailed). The 95% -confidence interval for this difference (henceforth, 95%-CID) is [1.6, 3.0] on the 11-point scale, and the effect is large according to Cohen's $d = .97$. Perhaps surprisingly, and not in line with N'gbala and Branscombe's (1997) results, the difference between regret attributed to the actor by participants in the *Actor Only* condition ($m = 5.1$) and regret attributed to the non-actor by participants in the *Non-actor* condition ($m = 4.2$) is marginally significant, $t = 1.94$, $p = .054$ (two-tailed), the 95%-CID being [0, 1.8] in favour of the actor. However, the effect size is small, CohenŠs $d = .35$.

Discussion

As expected, we obtain an action effect (and even a large action effect) using a within-subject design (the *Both Agents* condition). Contrary to expectations, we obtain a marginally significant action effect using a between-subject design; nonetheless, in addition to being only marginally significant, that effect is of a small magnitude, much smaller than it is using a within-subject design. We thus replicate, by and large, N'gbala and Branscombe's (1997) finding. The action effect seems to disappear using a between-subject design.

Two hypotheses are possible at this stage. It could be that the regret is almost the same for the actor and the non-actor (as it appears using the between-subject design), and that this slight difference is exaggerated in the within-subject design. But, as we will argue in the next section, it could also be that the straightforward use of continuous scales in a between-

subject design is an inappropriate tool to investigate differences in regret.

The Questionable Reliability of Regret Scales in Between-Subject Design

Ideally, we would like ratings to depend only on the stimulus being rated, and not on the person doing the rating or on the context of the rating. We would like ratings to be *measurements*. Rating the height of a person on a scale of 0 to 8 feet qualifies as measurement: Regardless of who is making the rating, we would still obtain the same score. One could be standing next to Michael Jordan and still obtain the same score. But is that true if we replace the 0 to 8 feet scale with an 11-point scale ranging from *very short* to *very tall*? Each judge would have her own idea of what it means to be "very short", and her own interpretation of which height would correspond to the third point of the scale, or to the sixth. Not that it would impinge on a judge's ability to use the scale to make paired-comparisons: If Melissa is perceivably taller than Julie, Melissa will get a greater rating on the 11-point scale than will Julie, whatever the interpretation of the scale the judge may hold.

Difficulties arise when judge A gives Melissa a rating on the 11-point scale (without having seen Julie), while judge B gives Julie a rating on the same 11-point scale (without having seen Melissa). As the two judges have their own interpretation of the scale, and do not necessarily share a common reference point, *any result can obtain*. The two women may receive the same rating. Melissa (the taller one) could even receive a lower rating than Julie.

It is quite clear that having two separate groups of judges rating the height of two persons on an 11-point scale from *very short* to *very tall* would not reliably tell us which one is taller. Hence, why would having two separate groups of judges rating the regret felt by two persons on an 11-point scale from *no regret at all* to *much regret* tell us reliably which one feels more regret?

Our point is that the use of scales in between-subject design may yield any result, regardless of the actual difference in perceived regret. Results of our Experiment 1 hinted at a marginally significant action effect, contrary to what N'gbala and Branscombe (1997) obtained. At least, the results were in the usual direction Ű but it is not unconceivable that we may have obtained an opposite effect, with inaction being regretted more than action. And it is noteworthy that a related phenomenon obtained in N'gbala and Branscombe's (1997) set of experiments.

Remember that in N'gbala and Branscombe's (1997) first study, Paul considered switching stocks, decided against it, and found out that he would have been better off by $1,200 if he had. Now in N'gbala and Branscombe's (1997) second study, another Paul considered going to see a game, decided against it, stayed home, and found later that his car had been wrecked by an unidentified driver in the parking lot, causing $1200 in damage which he had to pay for out of his own pocket. The financial loss is the same, the nature of the decision is the same (both Pauls eventually decided not to act), so which Paul should feel the more regret?

Assuming that regret is amplified when a decision is perceived as instrumental for the outcome (N'gbala & Branscombe, 1997) or when a decision-maker is perceived as responsible for the outcome (Zeelenberg et al., 2002), Paul 1 (the one making a bad investment decision) should feel more regret than Paul 2 (the one not going to see a game), as both his responsibility in the outcome and the instrumentality of his decision appear greater than do Paul 2's. Perhaps this is too strong a prediction, in which case one would expect the two Pauls to be attributed the same regret. But one would certainly not expect Paul 2 to be attributed more regret than Paul 1. Nevertheless, such is the case in N'gbala and Branscombe's (1997) results, with a mean regret of 68.00 for Paul 1 and a mean regret of 80.25 for Paul 2.

We have argued that one cannot draw reliable conclusions from regret ratings collected in between-subject designs. In order to further establish this point, we want to show in Experiment 2 that a between-subject design can fail to yield different regret ratings even when they clearly should be different.

The difference in regret after action and after inaction may not be especially intuitive. However, it is uncontroversial that, all other things being equal, a missed opportunity of $2,000 should lead to more regret than a missed opportunity of $1,000. And we certainly expect this difference to show up in a within-subject design. Yet will it show up in a between-subject design? Experiment 2 was conducted to answer this question.

Experiment 2

Method

Participants. A total of 98 undergraduate students, all native Chinese speakers, took part in the experiment (21 men and 77 women; mean age = 21.3, *SD* = 1.4).

Materials and design. Participants in the *Both Agents* condition were given the following scenario:

> Danny owned shares in company A. During the past year he switched to stock in company B. He now finds out that he would have been better off by $1,000 if he had kept his stock in Company A. Jefferson owned shares in company A. During the past year he switched to stock in company B. He now finds out that he would have been better off by $2,000 if he had kept his stock in Company A.

Note that in this scenario both agents are actors and made the same decision. The only difference between them is that in so doing, one of the agents missed a greater opportunity than the other. The order the characters appeared in the story was counterbalanced across questionnaires. Participants had to rate the regret felt by Danny and Jefferson, on two separate 11-point scales, ranging from 0 ("no regret at all") to 10 ("much regret").

Table 2
Experiment 2: Means and Standard Deviations in Perceived Regret for the Agent Losing $1,000 and for the Agent losing $2,000

	Losing $1,000	Losing $2,000
Both Agents	5.2 (2.5)	6.8 (2.2)
Loss $1,000 Only	4.6 (2.7)	n/a
Loss $2,000 Only	n/a	4.5 (3.0)

Participants in the Miss $1,000 Only and the Miss $2,000 Only conditions were provided with the information about Danny and Jefferson respectively and they had to rate the regret they felt about their decisions, again on an 11-point scale.

Procedure. Questionnaires were administered during a class. The experiment was conducted in Chinese.

Results

Table 2 presents the regret attributed to the agent missing the gain of $1,000 and to the agent missing the gain of $2,000, in the three experimental groups. Unsurprisingly, participants in the *Both Agents* condition expect the agent who misses $2,000 to experience more regret ($m = 6.8$) than the agent who misses $1,000 ($m = 5.2$) , $t = 4.98$, $p < .001$ (one-tailed). The 95%-CID is [.9, 2.3], and this effect is large, Cohen's $d = 1.0$. However, there is no reliable difference between regret attributed to the agent missing the $1,000 in the *Miss $1,000 Only* condition ($m = 4.6$) and regret attributed to the agent missing the $2,000 in the *Miss $2,000 Only* condition ($m = 4.5$), $t = .23$, *ns.*, the 95%-CID being [-1.2, 1.5].

Discussion

While the difference in regret after an action or a failure to act is not an especially intuitive one, the difference in regret after a missed opportunity of $2,000 and a missed opportunity of $1,000 is commonsense. Indeed, there appears to be a large difference in perceived regret when such regret is rated in a within-subject design. However, this difference disappears using the same rating scales in a between-subject design.

These results provide additional support to our suggestion that the straightforward use of regret scales in a between-subject design is an inappropriate tool to detect differences in perceived regret. However, we will now describe a technique that will improve the reliability of between-subject designs.

Improving the Reliability of Between-Subject Designs: The Common Reference Method

The problem we face is the following: Let us consider that a given participant rates the regret felt by a character in a situation A, and that another participant rates the regret felt by a character in a situation B. Those two ratings are not, we have argued, directly comparable. Yet, is it still possible to infer from those ratings which character is perceived as feeling the more regret?

We believe it is, provided that those absolute regret ratings are turned into relative regret scores, with the help of a third situation C, whose purpose is to serve as a common reference point to anchor the judgements.

Let us consider that the first participant rates the regret felt by a character in the situation C, and the regret felt by a character in the situation A. The second participant rates the regret felt by a character in the situation C, and the regret felt by a character in the situation B. What we are interested in is the difference between the two ratings given by the first participant, on the one hand, and the *difference* between the two ratings given by the second participant, on the other hand. If the gap between rating A and rating C is greater than the gap between rating B and rating C, then we will infer that the character in A is perceived as feeling more regret than the character in B.

The common reference method is meant to solve the problem with between-subject designs we have highlighted in this article. The fact that different participants may hold different interpretations of the scale is no longer problematic if the experimenter compares relative regret scores (relative to the common reference situation) rather than the absolute regret ratings provided by the participants.

Experiment 3 is a demonstration of the common reference method with respect to the action effect. We predict that an action effect will obtain with respect to relative regret scores, whatever the relation that obtains between absolute regret ratings.

Experiment 3

Method

Participants. A total of 125 undergraduate students, all native Chinese speakers, took part in the experiment (41 men and 87 women; mean age = 20.5, *SD* = 1.0).

Materials and design. We constructed two versions of the stock scenario, one containing the action, one containing the failure to act and both containing the same third event for reference. In the following *Common vs. Inaction* version, Paul is the target character and he is failing to act, while AlexŠs story serves as a reference point.

> Alex owned shares in company C. During the past year he switched to stock in company A. He now finds out that he would have been better off by $1,200 if he had switched his stock to Company B. Paul owns shares in company A. During the past year he considered switching to stock in company B, but he decided against it. He now finds out that he would have been better off by $1,200 if he had switched to the stock of company B.

Participants in the Common vs. Action condition were given the following scenario:

> Alex owned shares in company C. During the past year he switched to stock in company A. He now finds out that he would have been better off by $1,200 if he had switched his stock to Company B. George owned shares in company B. During the past year he switched to stock in company A. He now finds out that he would have been better off by $1,200 if he had kept his stock in company B.

In this second scenario, Alex is still making the same decision with the same consequences, and George is acting.

In both conditions, the order of appearance for the characters was counterbalanced across questionnaires. Participants had to rate the regret felt by Alex and Paul, Alex and George, on two separate 11-point scales, ranging from 0 ("no regret at all") to 10 ("much regret").

Procedure. Questionnaires were administered during a class. The experiment was conducted in Chinese.

Results and Discussion

Table 3 presents the regret attributed to the common agent, to the actor, and to the non-actor, in the two experimental groups. Regret attributed to the common agent is roughly similar (4.5 vs. 4.3) in the two conditions, $t = .36$, *ns.*, with a 95%-CID of [-1.0, .7].

Regret attributed to the actor in the *Common vs. Action* condition and regret attributed to the non-actor the *Common vs. Inaction* condition are not, as we have already pointed out, our measures of interest. In order to make use of the reference point we have introduced, we have to compare the difference between the regret of the actor and the regret of the common agent in the *Common vs. Action* condition (henceforth, the *actor's difference*), to the difference between the regret of the non-actor and the regret of the common agent in the Common vs. Inaction condition (henceforth, the *non-actor's difference*). Regret attributed to the common agent thus serves as a zero-point in comparing the regret of the actor and the regret of the non-actor. The results show that the actor's difference ($m = 3.43$, $SD = 2.38$, $N = 68$) is reliably higher than the non-actor's difference ($m = 1.68$, $SD = 3.59$, $N = 60$), $t = 3.19$, $p < .001$ (one tailed). The 95%-CID between the two scores is [.7, 2.8], and the effect is of medium size, Cohen's $d = .57$. As it was expected, the introduction of a common reference point, together with the transformation of regret ratings into relative regret scores, led to a recovery of the action effect in a between-subject design.

One concern about the results of Experiment 3 is whether they might be explained in terms of action-inaction fit. The common scenario in Experiment 3 featured an actor; thus, in the Common vs. Action condition, both characters were acting, whereas one character acted and one did not in the Common vs. Inaction condition. In order to control for this confounding, we applied the common reference method to the material of Experiment 2, where all scenarios featured actions. We now report the results of this test.

Experiment 4

Method

Participants. Fifty-seven undergraduate students, all native Chinese speakers, took part in the experiment (17 men and 40 women; mean age = 20.5, *SD* = .85).

Materials and design. Experiment 2 used two scenarios, one in which an agent missed a $2,000 opportunity, and one in which an agent missed a $1,000 opportunity. In line with the common reference method, we asked two groups of participant to rate, on an 11-point scale, the regret felt by one of these agents, and, on a separate 11-point scale, the regret felt by another agent missing an $800 opportunity. Thus, in the *Common vs. $2,000* version, participants could read that:

> Danny owned shares in company A. During the past year he switched to stock in company B. He now finds out that he would have been better off by $2,000 if he had kept his stock in Company A. Alex owned shares in company A. During the past year he switched to stock in company B. He now finds out that he would have been better off by $800 if he had kept his stock in Company A.

Participants in the *Common vs. $1,000* condition could read a similar scenario except that the Danny character missed a $1,000 opportunity rather than a $2,000 opportunity. In both conditions, the order of appearance for the characters was counterbalanced across questionnaires.

Procedure. Questionnaires were administered during a class. The experiment was conducted in Chinese.

Results and Discussion

We expected that, whatever the relation between the absolute regret ratings for the $2,000 miss and the $1,000 miss, the relative regret score for the $2,000 miss would be reliably higher than the relative regret score for the $1,000 miss. See Table 4 for the mean regret attributed to each character in each experimental condition. The mean regret rating for the $2,000 miss was 6.2 (*SD* = 2.8) and the mean regret rating for the $1,000 miss was 6.4 (*SD* = 2.7), $t = -.23$, *ns*, with a 95%-CID of [-1.6, 1.3]. As expected, however, the mean relative regret score for the $2,000 miss was 2.0, whilst the relative regret score for the $1,000 miss was only .9, $t = 2.78$, $p < .01$ (one-tailed), almost a large effect size ($d = .75$), with a 95%-CID of [.3, 1.9]. This demonstration further supports the reliability of the common reference method, as it allows observing, in a between-subject design, the commonsense difference between a $2,000 miss and a $1,000 miss, and rules out an explanation of Experiment 3's results in terms of action-inaction fit.

General Discussion

Since the publication of N'gbala and Branscombe (1997), a number of researchers (Anderson, 2003; Zeelenberg, van

Table 3
Experiment 3: Means and Standard Deviations of Perceived Regret for the Common Agent, the Actor, and the Non-actor

	Common Agent	Actor	Non-actor
Common vs Inaction (n=60)	4.5 (2.8)	n/a	6.1 (2.9)
Common vs Action (n=68)	4.3 (2.2)	7.7 (2.3)	n/a

Table 4
Experiment 4: Means and Standard Deviations in Perceived Regret for the Common Actor (Missing $800), the Actor Missing $2,000, and the Actor Missing $1,000

	Common Agent	$2,000	$1,000
Common vs $2,000 (N = 30)	4.2 (2.9)	6.2 (2.8)	n/a
Common vs $1,000 (N = 27)	5.5 (2.5)	n/a	6.4 (2.7)

der Plight, & de Vries, 2000) have stressed out methodological issues in the measurement of regret, and have shown specific concern regarding the challenge posed by the use of between-subject designs (see, e.g., Ordonez & Connolly, 2000; Zeelenberg, van Dijk, & Manstead, 1998, 2000. Our purpose in this article has been to investigate those issues and to provide a methodological solution to that challenge.

We have argued that absolute regret ratings collected in a between-subject design were not a reliable cue to underlying differences in perceived regret, for such ratings were not directly comparable. To overcome that difficulty, we have suggested the use of relative regret scores, computed with the help of a common reference situation.

Our common reference method still has to be refined, and its boundary conditions have to be examined. In this final section, we will consider in turn the three following issues: (a) Does the choice of the common reference situation matter? (b) Does our method make too bold an assumption vis-à-vis the metric properties of regret ratings? and (c) Can our method be applied to other domains than regret studies?

The Choice of the Common Reference Situation

Not any situation can be used as a common reference, as it is at least necessary to avoid both ends of the regret scale. That is, the common reference situation should be chosen so that it the regret it evokes is neither nonexistent nor too extreme.

Using as a common reference a situation that does not evoke any regret at all would amount to not using any common reference at all. On the other hand, using as a common reference a situation that evokes extreme regret would not be very helpful, as the right end of the regret scale lacks sensitivity. The latest point on the regret scale is not strictly speaking a point interval, as it stands for any feeling of very intense and painful regret. Hence, it is way too imprecise to be used in the computation of a relative regret score.

Once those minimal constraints are met, the choice of a specific common reference situation should be made freely, as it should not be of consequence to the final relative regret scores. Whether different common reference situations can lead to different experimental results may however be a topic for future, systematic research, that is beyond the scope of this article.

The Metrics of Regret Ratings

It could be that, in computing and comparing relative regret scores, we are making some unwarranted assumption on the metric properties of the regret scale. The fact that we use, as our measure of interest, the difference between two absolute regret ratings, implies that we treat the regret scale as an *interval* scale.

We have no argument to present in support of the view that the regret scale is indeed an interval scale. However, this point of view is implicit in all studies that present means and related inferential statistics based on regret scales. Thus, it can be said in defence of our method that however bold its assumptions on the metrics of the regret scale, it does not at least make any further assumption than previous regret research has already done (to the exception, of course, of research involving straightforward paired-comparisons, e.g., Gilovich & Medvec, 1994, or Zeelenberg et al., 2002).

Beyond Regret

While the focus of this article was on the psychology of regret (and more specifically on the action effect), other domains can benefit from the common reference method. The method can be used in any vignette study where participants have to rate the degree to which a character experiences a given emotion (satisfaction, happiness, anger, sadness, shame, etc.). Aside from emotions, the method can be used in conjunction with a vast number of scales, from responsibility scales to bad luck scales.

More generally, the common reference method is to be used as a way out of the *ill-defined scale dilemma*. When a researcher has to use as a dependent variable a rating on a scale whose interpretation may vary from one individual to another, he or she is faced with a dilemma. On the one hand, within-subject designs may lead to exaggerate the importance of what is actually a small difference. On the other hand, between-subject designs have every chance to yield unreliable results.

As a solution to this dilemma, we propose the use of the common reference method, which turns absolute ratings into relative scores. Hopefully, this suggestion will contribute to

the disentangling of significant experimental effects from artificial by-products of experimental designs.

References

Anderson, C. J. (2003). The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin*, *129*, 139–167.

Byrne, R. M. J., & McEleney, A. (2000). Counterfactual thinking about actions and failures to act. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1318–1331.

Connolly, T., Ordonez, L. D., & Coughlan, R. (1997). Regret and responsibility in the evaluation of decision outcomes. *Organizational Behavior and Human Decision Processes*, *70*, 73–85.

Gilovich, T., & Medvec, V. H. (1994). The temporal pattern to the experience of regret. *Journal of Personality and Social Psychology*, *67*, 357–365.

Gilovich, T., Medvec, V. H., & Chen, S. (1995). Omission, commission, and dissonance reduction: Overcoming regret in the Monty Hall problem. *Personality and Social Psychology Bulletin*, *16*, 284–295.

Gleicher, F., Kost, K. A., Baker, S. M., Strathman, A. J., Richman, S. A., & Sherman, S. J. (1990). The role of counterfactual thinking in judgments of affect. *Personality and Social Psychology Bulletin*, *16*, 284–295.

Inman, J. J., & Zeelenberg, M. (2002). Repeat versus switch decisions: The attenuation role of decision justifiability. *Journal of Consumer Research*, *29*, 116–128.

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, *246*, 160–173.

Landman, J. (1987). Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin*, *13*, 524–536.

N'gbala, A., & Branscombe, N. R. (1997). When does action elicit more regret than inaction and is counterfactual mutation the mediator of this effect? *Journal of Experimental Social Psychology*, *33*, 324–343.

Ordonez, L. D., & Connolly, T. (2000). Regret and responsibility : A reply to Zeelenberg et al. (1998). *Organizational Behavior and Human Decision Processes*, *81*, 132–142.

Seta, J. J., McElroy, T., & Seta, C. E. (2001). To do or not to do: Desirability and consistency mediate judgments of regret. *Journal of Personality and Social Psychology*, *80*, 861–870.

Zeelenberg, M., van den Bos, K., van Dijk, W. W., & Pieters, R. (2002). The inaction effect in the psychology of regret. *Journal of Personality and Social Psychology*, *82*, 314–327.

Zeelenberg, M., van der Plight, W. W., & de Vries, N. K. (2000). Attributions of responsibility and affective reactions to decision outcomes. *Acta Psychologica*, *104*, 303–315.

Zeelenberg, M., van Dijk, W. W., & Manstead, A. S. R. (1998). Reconsidering the relation between regret and responsibility. *Organizational Behavior and Human Decision Processes*, *74*, 254–272.

Zeelenberg, M., van Dijk, W. W., & Manstead, A. S. R. (2000). Regret and responsibility resolved? Evaluating Ordóñez and Connolly's (2000) conclusions. *Organizational Behavior and Human Decision Processes*, *81*, 143–154.