Abstract

It has been established that social processes play an important role in achieving and maintaining a healthy lifestyle, but there are still gaps in the knowledge on how to apply such processes in behavior change interventions. One of these mechanisms is social comparison, i.e. the tendency to self-evaluate by comparing oneself to others. Social comparison can be either downward or upward, depending on whether individuals compare themselves to a target that performs worse or better. Depending on personal preferences, the variants can have beneficial or adverse effects. In this paper, we present the results of an experiment where participants (who indicated to prefer either upward comparison or downward comparison) were sequentially shown both directions of social comparison, in order to influence their physical activity levels. The results show that presenting users with the type of social comparison they do not prefer may indeed be counter-effective. Therefore, it is important to take this risk into account when designing physical activity promotion programs with social comparison features.
This chapter appeared as:

12.1 Introduction

Physical inactivity is a major public health issue. Evidence shows that it increases the risk of many health problems, including non-communicable diseases such as type-2 diabetes, cardiovascular diseases, cancer and mental illnesses (Lee et al., 2012). Consequently, physical inactivity is the fourth leading cause of death globally (Kohl et al., 2012). Vice versa, engaging in sufficient levels of physical activity has been associated with effects on both physical and mental health (Conn et al., 2011; Eime et al., 2013; Reiner et al., 2013). However, a large part of the Western population does not meet the global recommendations of being moderately to vigorously active for at least 30 minutes per day on at least five days per week (Hallal et al., 2012). Therefore, healthy lifestyle promotion programs and behavior change interventions are a priority in most Western countries, in order to increase global physical activity levels.

It has been established that social processes play an important role in achieving and maintaining a healthy lifestyle (Zimmerman and Connor, 1989). Several mechanisms that underlie these social influences have been identified, such as priming, social norms, behavior modeling, social facilitation and social support (Bandura, 1998; Cheng et al., 2014; McNeill et al., 2006), but there are still gaps in the knowledge on how to apply such processes in human behavior change interventions. Another mechanism that plays a part in behavior evaluation and behavior change is social comparison (Buunk et al., 2013; Festinger, 1954), which is often applied in physical activity promotion apps as a form of providing feedback to the user (Middelweerd et al., 2014). Social comparison exists in two variants: downward social comparison and upward social comparison (Festinger, 1954), depending on whether the target (with whom one compares oneself) performs worse (i.e., downward) or better (i.e., upward) than the individual. Both variants can be effective and encouraging, for instance by boosting one’s self-view or by motivating improvement, but also counter-effective and discouraging, for instance by advocating inferior standards or by threatening the self-view (Corcoran et al., 2011). Therefore, it is important to carefully design interventions that incorporate a social comparison component.

In this paper, we describe an experiment to test whether social comparison of physical activity via an online intervention leads to a measurable effect on the behavior of participants. More specifically, we investigate whether the direction of the presented social comparison (upward or downward) indeed yields two-sided effects on people’s physical activity levels, depending on the users’ indicated preference for one of these two variants. At the same time, the results of the experiment will indicate whether people are able to assess their own preference truthfully and effectively, according to the effects on their behavior.

The remainder of the paper is organized as follows: Section 12.2 describes some background on social comparison theory. In Section 12.3, we describe the methods used to gather and analyze the data. The results are presented in Section 12.4, and reflected upon in Section 12.5. Finally, Section 12.6 closes the paper with a conclusion.

12.2 Background

Social comparison is defined as the tendency to evaluate oneself through comparison to others, which is an important source of competitive behavior to self-improve (Garcia et al., 2013). In both upward and downward social comparison, people aim to attain or maintain a higher level of performance than others (Festinger, 1954). The desire to achieve or keep
such a superior position is called a ‘comparison concern’ (Garcia et al., 2013). The model presented in (Garcia et al., 2013) shows that two sets of factors can encourage competitive behavior by raising such comparison concerns: individual and situational factors.

Individual factors are those that may vary greatly between individuals, even if they find themselves in comparable situations. The three most important individual factors are relevance, similarity and closeness (Goethals and Darley, 1977). The more relevant a particular dimension of performance (e.g., income, study results or sports achievements) is to an individual, the stronger their comparison concerns will be (Hoffman et al., 1954). Likewise, the more similar a target with whom one compares oneself is, the stronger the effect of the social comparison will be (Kilduff et al., 2010). Finally, the comparison concerns are also stronger when the individual and the target have a close personal relationship than when they don’t know each other (well) (Tesser and Smith, 1980).

Situational factors are those that concern an individual’s perception of the surrounding social environment, by which means they can yield a more general effect on similarly situated individuals (Garcia et al., 2013). Several situational factors can contribute to one’s comparison concerns. For example, incentive structures (i.e., the incentives associated with the comparison) can encourage competitiveness when higher values are expected in case of better (relative) performance. Another factor is the proximity to a standard, i.e. whether an individual is close to the number-one position (or some other meaningful performance metric). The closer to such a standard, the stronger the comparison concerns are. The number of competitors also influences comparison concerns: the fewer competitors, the stronger the competitive behavior of individuals. A final example of a situational factor in social comparison is social category fault lines: when comparing to targets from other social categories (based on gender, nationality, etc.), the comparison concerns are stronger than when comparing within such categories.

As mentioned in the Introduction, social comparison can involve a target (with whom one compares oneself) that performs better (i.e., upward) or worse (i.e., downward) than the individual. The two variants address different underlying motivational processes, implying different benefits and drawbacks. The main positive effect of downward social comparison is that comparing to a lower-performing target can boost the individual’s self-esteem and subjective well-being (Wills, 1981). On the other hand, downward comparison could also result in relatively low goals, since it doesn’t challenge an individual to try harder to minimize the discrepancy with someone else’s performance. The main benefit of upward social comparison is exactly that: it motivates people to self-improve in order to approximate better performing others (Lockwood and Kunda, 1997). In addition, a higher-performing target could provide information and serve as a role model, which allows an individual to learn how to perform better (Maddux, 1995). However, if one deems the performance of superior others to be unattainable, upward social comparison could have a discouraging and deteriorating effect as well (Lockwood and Kunda, 1997).

The preference for either upward or downward social comparison can originate from the motivation behind the social comparison. Although self-evaluation can be achieved with both upward and downward comparison, individuals who strive for self-enhancement (i.e., boosting their self-view) are more inclined towards downward comparison (Suls et al., 2002; Wills, 1981). Similarly, people who pursue self-improvement will tend to opt for upward comparison (Suls et al., 2002). Research has shown that men more often engage in upward social comparison, whereas women tend to compare in a downward direction.
Based on the theory about social comparison described above, an experiment has been set up. The details of the setup and the data collection are described below.

12.3.1 Social comparison design and implementation

The social comparison intervention was implemented as a dynamic web application, developed by Werkhooven (2015), in which users could see their own activity data and a comparison with their friend’s data. The web application was accessible through a web browser, and the information displayed in the web application was therefore available to the users at any time and at any location. The activity data was registered using a wireless activity monitor, the Fitbit One (Fitbit Inc., no date), which tracks steps, floors climbed, distance, calories burned and active minutes. For this experiment, the number of steps was taken as main measure for physical activity.

The web application first shows an overview of all participating users, with a name and a profile picture. Then, the users can log in to their personal page by clicking on their own picture or name. The personal pages are secured with a user password, in order to protect the privacy-sensitive information, but also to prevent users from looking at each other’s personal pages.

The personal page in the web application consists of three parts. First, a profile picture and the name of the user are shown at the top of the page. Second, the page contains a graph where the user’s own activity data is visualized. The graph shows the number of steps for the past 24 days in a bar chart. Third, the personal page shows a comparison of the user’s activity data with the activity data of two friends. This comparison is established through a line chart with three lines, a thick solid blue line for the user’s own activity data and thinner solid red and yellow lines for the two friends. The area below the lines was colored in the same color as the lines. Figure 12.1 shows an example of a personal page on a mobile phone.

The choice for which other users’ activity data to show to the user depends on the user’s preference (upward or downward) and the current setting of the system (in line with the user preference or the opposite of the user preference). When the web application employs downward social comparison, the average number of steps of the friends is lower than the user’s average number of steps. Similarly, when the web application shows upward social comparison, the friends have a higher average number of steps than the user. This implies that the physical activity levels of friends were higher or lower for the overall period, but that activity levels on a specific day could be opposite. In other words, the direction of the social comparison is relatively subtle. Another consequence of this setup was that the friends that were shown in the graphs could change per day.

12.3.2 Experimental setup

The effect of the social comparison intervention was tested in a small user study. Twenty members of a Dutch amateur soccer club participated. The participants were between 21 and 26 years old, with an average age of 24.6 years (sd = 1.2 years). The participants were all male, and the vast majority of the group played soccer on the same team. Although the
homogeneity of the group evidently limits the possible conclusions that can be drawn from this experiment (see Discussion in Section 12.5), this was a deliberate choice in order to reduce the differences in individual and situational factors (see Background in Section 12.2)
as much as possible.

All participants were asked to use an activity monitor, the Fitbit One™, for a period of four weeks. This four-week experiment period was preceded by a test phase of five days, in which the participants could get used to wearing the device and monitoring their own physical activity. That way, the Hawthorne effect (i.e., the effect of knowing that one is being observed (Parsons, 1974; Sommer, 1968)) and the novelty effect (i.e., the effect of using a new gadget that stimulates physical activity) could hopefully be diminished. The four-week experiment period was divided into two parts: in the first two weeks, ten randomly chosen participants received upward social comparison and the other ten participants received downward social comparison, and vice versa in the last two weeks. This implies that our experiment adopts a partial within-subjects approach, allowing us to test the effect of both directions on each individual participant. During the four weeks that the participants received either one of the forms of social comparison, they were asked to view their personal page on a daily basis. Table 12.1 shows which group received which direction of social comparison in which period.

Table 12.1: Direction of social comparison per group and period.

<table>
<thead>
<tr>
<th></th>
<th>Period 1 (week 1–2)</th>
<th>Period 2 (week 3–4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Upward</td>
<td>Downward</td>
</tr>
<tr>
<td>Group 2</td>
<td>Downward</td>
<td>Upward</td>
</tr>
</tbody>
</table>

At the end of the four weeks, the participants were asked to fill out a short questionnaire regarding their social comparison preferences. The questionnaire was distributed to the participants after the experimental period, in order to limit their knowledge about the aim of the experiment. The questionnaire consisted of five questions, as shown in Figure 12.4.
Chapter 12. Effectiveness of social comparison on physical activity

The participant’s indicated preference was determined based on question Q2. Using the information about the preference and the setup shown in Table 12.1, it could be determined when the participants were shown social comparison according to their own preference (period 1 or period 2), and when they were shown social comparison opposed to their preference.

Q1. Did you enjoy to see your physical activity compared to other users? (Five-point Likert scale from ‘not at all’ to ‘very much’.)
Q2. Do you prefer to compare your physical activity to individuals that perform worse or individuals that perform better? (Dichotomous: ‘individuals that perform worse’ or ‘individuals that perform better’.)
Q3. How often do you try to figure out what people think who encounter the same problems as you? (Five-point Likert scale from ‘almost never’ to ‘almost always’.)
Q4. How often does it make you feel good when coworkers or classmates perform worse than you? (Five-point Likert scale from ‘almost never’ to ‘almost always’.)
Q5. How often does it make you feel challenged when coworkers or classmates perform better than you? (Five-point Likert scale from ‘almost never’ to ‘almost always’.)

Figure 12.4: Questionnaire about social comparison preferences.

12.3.3 Analysis

The experimental setup described in the previous two subsections combines three different dimension. First, the participants are exposed to either upward or downward social comparison on their personal page. Second, the participants are exposed to the direction of social comparison that is either in line with their preference or opposed to their preference. Third, the participants receive a certain direction of social comparison in the first two weeks of the experiment (period 1) or in the last two weeks (period 2). This setup allows for analysis of the data from different angles.

In accordance with the social comparison theory, we expect that participants who receive social comparison in line with their preference will show an increase in physical activity, whereas showing social comparison opposed to their preference will have adverse effects. Therefore, the following hypotheses were formulated:

H1. Participants who prefer upward social comparison and receive upward social comparison will show an increase in physical activity.
H2. Participants who prefer upward social comparison but receive downward social comparison will show a decrease in physical activity.
H3. Participants who prefer downward social comparison and receive downward social comparison will show an increase in physical activity.
H4. Participants who prefer downward social comparison but receive upward social comparison will show a decrease in physical activity.
H5. Participants who receive social comparison in line with their preference will show an increase in physical activity, and participants who receive
social comparison opposite to their preference will show a decrease in physical activity.

The effect of each social comparison condition is investigated with a Mann-Kendall test, a non-parametric test for statistical dependence. Kendall’s $\tau$ (tau) indicates to what extent the orderings of the data are similar when ranked by each of the variables.

12.4 Results
In this section, the results of the experiment are presented. In Section 12.4.1, the participants’ preferred directions of social comparison are summarized. Then, Section 12.4.2 shows a first glance at the results. The following subsections present the data in more detail: in Section 12.4.3 to Section 12.4.7, we show the results associated with hypothesis H1 to hypothesis H5, respectively.

12.4.1 Social comparison preferences
The experiment described in the previous section was performed with 20 participants. Based on the questionnaire depicted in Figure 12.4, in particular question Q2, it was determined that fifteen participants preferred upward comparison and five participants preferred downward comparison.

12.4.2 Exploratory data analysis
The first question that comes to mind when designing and executing such a relatively small-scale experiment, is whether the sample size and duration are substantial enough to see effects of the social comparison intervention on the physical activity behavior. Figure 12.5 shows the average number of steps of the nine participants who received the direction of social comparison that they did not prefer in the first two weeks and the direction that they preferred in the last two weeks. It is clear to see that the daily step counts decrease substantially in the first two weeks, and after switching to their preferred direction of social comparison, the daily step counts start to increase again. This suggests that even though the experiment is relatively small, the results provide sufficient basis for further analysis.

In the next subsections, we present the data gathered in the experiment in more detail, according to the hypotheses introduced in Section 12.3.3.

12.4.3 Hypothesis H1: Preferred upward, presented upward social comparison
This section presents the data of the participants who preferred and received upward social comparison. Figure 12.6 shows the average number of steps of these 15 participants over a period of two weeks.

The Mann-Kendall test on this data yields that $\tau = .385$, which indicates a moderate positive rank correlation. However, the rank correlation is just under significant, at a p-value of $p = .0617$. The results are summarized in Table 12.2.

12.4.4 Hypothesis H2: Preferred upward, presented downward social comparison
This section shows the data of the participants who preferred upward social comparison but received downward social comparison. Figure 12.7 shows the average number of steps of these 15 participants over a period of two weeks.
Chapter 12. Effectiveness of social comparison on physical activity

Figure 12.5: Average number of steps in four-week period (with linear trend line), for participants who received social comparison opposite to their preference in the first two weeks and social comparison in line with their preference in the last two weeks.

Figure 12.6: Average number of steps in two-week period, for participants who preferred and received upward comparison.

The Mann-Kendall test on this data yields that $\tau = -.538$, which indicates a considerable negative rank correlation. The rank correlation is also significant at a $p$-value of $p = .0067$. The results are again summarized in Table 12.2.

12.4.5 Hypothesis H3: Preferred downward, presented downward social comparison

This section shows the effect of downward social comparison for participants who also preferred this direction. Figure 12.8 shows the average number of steps of these 5 participants over a period of two weeks.

The Mann-Kendall test on this data yields that $\tau = -.297$, which indicates a fair negative rank correlation. However, the rank correlation is not significant at a $p$-value of $p = .157$. Again, the results are summarized in Table 12.2.
12.4 Results

Figure 12.7: Average number of steps in two-week period, for participants who preferred upward comparison and received downward comparison.

Figure 12.8: Average number of steps in two-week period, for participants who preferred and received downward comparison.

12.4.6 Hypothesis H4: Preferred downward, presented upward social comparison

This section presents the effect of showing downward social comparison for participants who preferred to receive upward social comparison. Figure 12.9 shows the average number of steps of these 5 participants over a period of two weeks.

The Mann-Kendall test on this data yields that $\tau = -0.560$, which indicates a considerable negative rank correlation. The rank correlation is also significant at a $p$-value of $p = .0046$. The results are again summarized in Table 12.2.

12.4.7 Hypothesis H5: Preferred vs. non-preferred social comparison

In this section, the data is reorganized to two groups: participants who receive social comparison in line with their preference (“preferred”) and participants who receive social comparison opposite to their preference (“non-preferred”). Note that all participants are
Chapter 12. Effectiveness of social comparison on physical activity

Figure 12.9: Average number of steps in two-week period, for participants who preferred downward comparison and received upward comparison.

represented in both groups, as everyone received both directions of the social comparison in either the first two weeks or the last two weeks.

Figure 12.10 shows the average number of steps of the two weeks when the participants received their preferred social comparison, and Figure 12.11 shows the period when they received the opposite direction. Figure 12.10 shows a fair positive rank correlation ($\tau = .253$), which is not significant ($p = .233$), and Figure 12.11 shows a strong negative rank correlation ($\tau = .648$) that is also significant ($p < .001$).

Figure 12.10: Average number of steps in two-week period, for participants who received social comparison in line with their preference.
12.5 Discussion

The results presented in Section 12.4 show that even though the results are not always significant, several interesting trends are visible. Except for the findings associated with hypothesis H3, the direction of Kendall’s $\tau$ is always in line with the hypothesized results. Moreover, on all occasions when the participants are presented the version of social comparison they do not prefer, there is a moderate to strong negative rank correlation, which is also strongly significant. This indicates that user preferences are indeed important: if not to enhance the motivational effects of social comparison, then at least to avoid the adverse effects of showing users social comparison that discourages them. As social comparison features implemented in physical activity promotion programs frequently show some form of upward comparison (e.g., leaderboards, high scores), it is important to acknowledge the risk of such a representation on the performance of people who prefer the opposite direction of comparison. Often, when physical activity platforms incorporate a ranking among friends, the ranking shows other users with similar activity levels (e.g. three places ahead and three places behind), such as on the Fitbit dashboard webpage (Inc., 2016). Sometimes, the top performers are shown as well (i.e., upward comparison), as for example in the Human app (Human.co, 2016). When showing a certain type of social comparison by default that is

Table 12.2: Mann-Kendall’s rank correlation tests.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Kendall’s $\tau$</th>
<th>p-value</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>.385</td>
<td>.0617</td>
<td>-</td>
</tr>
<tr>
<td>H2</td>
<td>-.538</td>
<td>.0067</td>
<td>*</td>
</tr>
<tr>
<td>H3</td>
<td>-.297</td>
<td>.157</td>
<td>-</td>
</tr>
<tr>
<td>H4</td>
<td>-.560</td>
<td>.0046</td>
<td>*</td>
</tr>
<tr>
<td>H5a (preferred)</td>
<td>.253</td>
<td>.233</td>
<td>-</td>
</tr>
<tr>
<td>H5b (non-preferred)</td>
<td>-.648</td>
<td>&lt; .001</td>
<td>*</td>
</tr>
</tbody>
</table>
opposite to a user’s preference, such features could inadvertently lead to counter-effective outcomes.

When looking at the results in more detail, it can be seen that the findings show potential for stronger conclusions in bigger follow-up experiments. For example, after discarding the last day of the data associated with hypothesis H1 (see Figure 12.6 in Section 12.4.3), which was an usually “lazy” Sunday (note that Sundays usually have low average activity levels in the graphs shown above), the trend is immediately strong ($\tau = .590$) and significant ($p = .0043$). Similarly, when discarding the last day of the data associated with hypothesis H5a (see Figure 12.10 in Section 12.4.7), the trend becomes more considerable ($\tau = .424$) and closer to significance ($p = .063$).

When considering the significance of the results, the relative small sample size should be taken into account. After all, the results of hypotheses H1 and H2 were based on the data of only fifteen participants, and the results of hypotheses H3 and H4 even on only five participants. It is striking that still significant effects could be found based on such a small sample size. The small sample size could also explain some of the unexpected results. For example, this group of participants who preferred downward social comparison included a participant with highly varying step counts (e.g., < 1,000 and > 34,000), which could have distorted the group averages.

Indeed, the findings ask for a larger experiment to confirm the results and conclusions of the current study. Also, it would be interesting to see whether the results extend beyond the homogeneous participant population of this study, such as to women, non-exercisers, other age groups and users who do not know each other. Since the participants in this study were members of an amateur soccer club, it can be expected that they are relatively active and healthy. Therefore, it would be especially interesting to see whether these results transfer to a population of inactive people, as they should be the target group of physical activity promotion programs, and because they might experience social comparison differently.

In addition, the current study only investigated the effect of one particular implementation of social comparison. It would be interesting to explore other visualizations of social comparison (e.g., rankings, sorted charts), to see whether the specific representation influences the effect on the users. Another limitation of this current experiment is that the participants’ preference is based on one dichotomous question only. One can imagine that some people do not appreciate social comparison at all, but this was not an option in the current setup. Although the results indicate that people are reasonably able to assess their own preference, further research could investigate other (possibly implicit) ways to determine the user’s preferences. For example, instead of querying it directly, one could try to derive the preferred direction from a (personality) questionnaire, or from the user’s behavior when exposed to the two different variants. This could be helpful in particular for people who do not have a strong personal preference for one of the two options.

### 12.6 Conclusion

Social comparison features implemented in most physical activity promotion programs often show some form of upward comparison: e.g., leaderboards and high scores. However, the results of this experiment show that presenting users with such types of social comparison may have adverse effects for people who indicate that they prefer downward social comparison. Even though based on a small sample size, our results indicate that the adverse
effects are stronger and more significant than the positive effects. Therefore, it is important that intervention designers take this risk in account when developing a physical activity promotion program with social comparison features.

**Acknowledgments**

The authors would like to thank Kevin van Werkhooven for the development of the social comparison intervention and his contribution to the setup and execution of the experiment described in this paper.
Chapter 12. Effectiveness of social comparison on physical activity

References


