Published in final edited form as:
Nat Hum Behav. 2018 October ; 2(10): 778–787. doi:10.1038/s41562-018-0436-y.

Network Interventions for Changing Physical Activity Behaviour in Preadolescents

Antonios Proestakis[1,*], **Eugenia Polizzi di Sorrentino**[1,2], **Helen Elizabeth Brown**[3], **Esther van Sluijs**[3], **Ankur Mani**[4], **Sandra Caldeira**[1], and **Benedikt Herrmann**[1]

[1]Joint Research Centre, European Commission

[2]Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

[3]UKCRC Centre for Diet and Activity Research (CEDAR) and MRC Epidemiology Unit, University of Cambridge, Cambridge, UK

[4]Industrial and Systems Engineering, University of Minnesota

Abstract

Network interventions can help achieve behavioural change by inducing peer-pressure in the network. However, inducing peer-pressure without considering the structure of the existing social network may render the intervention ineffective or weaker. In a 7-week school-based field experiment using preadolescents' physical activity (PA) as a proxy for estimating behavioural change, we test the hypothesis that boys' and girls' distinct networks are susceptible to different social incentives. We run three different social-rewards schemes, in which classmates' rewards depend on the PA of two friends either reciprocally (directly or indirectly) or collectively. Compared to a random-rewards control, social-rewards schemes had an overall significantly positive effect on PA (51.8% increase), with females being more receptive to the direct reciprocity scheme (76.4%) and males to team (collective) rewards (131.5%). Differences in the sex-specific sub-networks can explain these findings. Network interventions adapted to the network-specific characteristics may constitute a powerful tool for behavioural change.

Peer to peer interactions can be an important source of social influence in the diffusion of ideas and behaviours in both physical and digital social networks1. The quality

---

*Correspondence to: Antonios.Proestakis@ec.europa.eu.

Data availability

The datasets generated and analysed during the current study are available from the corresponding author upon request.

Code Availability

The code (.do files) used in our statistical analysis and image production is available upon request from the corresponding author.

Author contributions

A.P., B.H., A.M., S.C. developed the concept; A.P., E.P.S., B.H. designed the experiments; A.P., B.H., E.P.S, S.C. obtained the ethical approval and the authorisation from data protection officer; E.P.S., S.C., A.P. recruited the students and took permissions from the schools' directors/teachers; E.P.S., A.P. performed the experiments, H.E.B and E.v.S. led on physical activity measurement and processing; A.P., E.P.S., A.M. analysed the data; B.H. supervised the study A.P., S.C., E.v.S., H.E.B, E.P.S, B.H., A.M. wrote the paper.

Competing interests

The authors declare no competing interests.

characteristics of interpersonal ties and interactions (e.g. strength, direction, frequency, locality, etc.) among the members of the network determine the magnitude of peer-pressure exerted by one peer on another and eventually the size of social influence in the network2. Network induction or alteration interventions3 can achieve behavioural change by changing or fixing some of these quality characteristics. However, different subgroups of individuals with different quality characteristics and/or conflicting interests may exist within the same network (sub-networks). This makes isolating the effect of different network interventions difficult4. In this study, we take the advantage of the controlled environment of sex-segregated school-class networks5 to induce different network interventions (based either on collective or reciprocal interaction schemes) and measure their impact in boys' and girls' distinct sub-networks6,7. As obesity and physical fitness have been established in the literature as "contagious" concepts8, 9, 10, 11 we use physical activity as a proxy for conceptualizing behaviour and estimating behavioural change. As changing usual behaviour and habits is a complex task2 requiring focused interventions, further generalisation of the results should however be done with caution.

Introducing social norms or inducing peer-pressure can both result in social influence. In the first case, individuals are not required to take any conscious action while contagion happens through observation and copying others' behaviour, perceived (or presented or induced in the context of network interventions) as social norms (descriptive or injunctive)12–14. In the second case, the focus of this study, peer-pressure requires individuals and peers to make conscious decisions to exert pressure15 as a response to a specific external stimulus which inter-relates peers' welfares. One can consider that also in this case, a new social norm is created and maintained at least for the period during which an external stimulus exists16,17. The power of social influence in various socio-economic areas is highlighted in several empirical studies (see review18). Recent field-, laboratory- and internet-based network experiments or interventions have demonstrated a role for peer-pressure in social mobilization for problem solving19, product adoption1, health behaviours20 or altruism21). Peer-pressure plays a particular important role in the promotion of cooperative behaviour22. In evolutionary biology, it is well established that cooperation evolves through different mechanisms23 which all work by exploiting the local-dyadic interactions (i.e. direct reciprocity, kin selection), the global-network interactions (i.e. indirect reciprocity) or both (i.e. network reciprocity, group selection). Social influence in general, and peer-pressure in particular, can be intentionally induced through carefully designed network interventions targeting socially desirable outcomes. The term "induction interventions" was introduced by Valente to classify interventions that stimulate or force peer-to-peer interactions to create cascades in information or behavioural diffusion3. Moreover, interventions on the network structure either by adding or deleting nodes and links, or by rewiring existing links are classified as "alteration interventions". The experiment presented here incorporates elements from both types of interventions. Building on the theoretical model by Mani and colleagues24 (and also its experimental application25), we induce peer-pressure by reallocating individuals' positive externalities towards their friends. The performance of an individual affects only and repeatedly his/her close-friends' payoffs. At the same time, we intervene in the structure of the network by suspending any rewiring process which would naturally emerge26, 27 as a reaction to the negative (i.e. lack of positive) externalities suffered by an individual due to his/her close-friends' adverse performance - a process which in evolutionary game theory is dubbed "spatial selection"15 or "link reciprocity"28. Exerting peer-pressure towards harmful (not beneficial) friends is the only alternative reaction to be taken by someone who is interested in improving his/her personal payoff. Physical activity is used in this study as a "behavioural proxy" for estimating the size of behavioural change after our interventions. However, it was also chosen for its importance on children’s physical and mental health29; developing and evaluating effective interventions to increase physical activity in children is a public health priority30 and one of the core points of our study. We approach the promotion of physical activity as a "complex contagion" problem2 which requires multiple exposures to a stimulus (our experimental incentives) reinforced by multiple strong-tied sources (participants' friends). Using peer-pressure induction and network-rewiring suspension as basic principles, we designed three different social-rewards schemes, based on team (collective)26, 31, direct reciprocal21 or indirect reciprocal32,33 interaction schemes and test them, together with the individual reward schem34, 35 against a random reward scheme which represents our control condition.

We conducted the study in two experimental phases (see methods), with 10 fifth-grade (i.e. 9-11 year-olds) elementary-school classes participating in the winter phase (92 females and 84 males from 7 schools) and 9 different classes in the spring phase (85 females and 88 males from 8 schools). Each class participated in only one of the experimental conditions while classes from the same school and phase were assigned to the same condition to avoid contamination. Following ethical and data protection approval (see Methods: Ethical approval), and parental consent, network-data was collected with the help of the class teachers who served as mediators between students and researchers throughout the whole experiment to preserve students' anonymity. For the same reason, researchers also avoided collection of any other data not directly linked to the intervention (e.g. socio-economic data). The strength of all dyadic ties ($1 \le \tau_{ij} \le 5$) between any two classmates i and j was elicited through a self-reported network questionnaire. This information was then used for the formation of interacting teams or pairs (consisting of same-sex relatively good friends ($\tau_{ij} \ge 3$ and $\tau_{ji} \ge 3$), henceforth partners) in the social-rewards schemes. Inducing same-sex groups was both fair, due to the known sex differences in physical activity36, and also natural, due to the existence of sex homophily in the actual networks (E-I-index($\tau_{ij=5}$)= -0.54, p_{value}<0.001, see 37 and "Supplementary Notes: Network Analysis" and supplementary fig. S7-S8).

In both winter and spring phases, following a one-week preliminary period (p_0), each school was randomly assigned to one of five study conditions and received the corresponding instructions about incentives, rewards and grouping information (where applicable). The interventions were conducted in parallel for five consecutive weeks divided into four periods: three one-week long periods (p_1,p_2 and p_4) and one longer period (p_3) of two weeks. Data was also recorded for an additional one-week long post-experimental period (p_5) in which students did not receive any kind of incentive or reward.

Physical activity was continuously recorded using Actigraph accelerometers (either GT1M, GT3X or GT3X+; ActiGraph, Pensacola, FL.) throughout the seven-week experiment (except for the days in which data had to be extracted and prepared for participants' feedback, see supplementary fig. S1). Using Evenson and colleagues'38 cut-points, minutes spent in moderate-to-vigorous physical activity (>2296 counts per minute (c.p.m.)) was communicated to participants as "minutes of intensive physical activity". This information was given to each participant individually (in a sealed envelope) at the onset of each period (for all $p_{t \geq 1}$) together with information on their absolute performance, their relative classification (described as Gold, Silver or Bronze medal) and the corresponding points (100, 80 or 60) obtained in the previous period (p_{t-1}). In the same letter, boys and girls were also informed about the sex-specific upper and lower thresholds for achieving one of the classification medals in p_t. These thresholds corresponded to the upper and lower tertiles of boys' and girls' physical activity distributions in p_{t-1} (see Methods: Physical Activity Measurement and Feedback). In the social-rewards schemes, participants were also informed about the relative performance and points of their partners.

The different experimental conditions assigned to the school classes defined the allocation rules of the points earned. In the individual reward scheme, participants received the points corresponding to their own relative performance. In the social rewards conditions, points were not dependent on their own but on partners' performance (fig. 1). In the direct and indirect reciprocity schemes each individual’s points were divided between two partners (point takers), while they themselves received half of the points from each one of the same (in direct reciprocity) or other (in indirect reciprocity scheme) two partners. Finally in the team scheme, three partners were assigned to the same team and engaged in a sharing scheme similar to a Public Goods Game39; the points of all team members are summed up and then shared equally among them. At the end of the experiment (during p_5), all points from all periods were added for each participant individually who could exchange them for one or more items from a gift-list. In the control condition, students entered a draw for having worn the accelerometer correctly. At the end of each experimental period (p_1-p_4), these participants randomly earned 100, 80 or 60 points (with 1/3 chance), irrespective of their physical activity performance. The expected payoff per period (for all $p_{t \geq 1}$) for a random student in all conditions, including the control, was equal to 80 points. The total expected payoff from all periods ($p_{t \geq 0}$) was 420 points (min=340 points, max=500 points).

Results

Following standard processing practices in child-based accelerometry40,41, data before 07:00 and after 23:00 were excluded and non-wear time was removed. An additional filter excluding days with less than 10 hours of valid wear time left 9776 valid days with a mean of 28.1 $CI_{95\%}$[13.4, 42.8] days per participant for our analysis (supplementary table S1). A minimum of three independent classes were assigned to each experimental condition with a mean of 35.4 females $CI_{95\%}$[19.3, 51.5] and 34.4 males $CI_{95\%}$[20.9, 47.9] participating in each experimental condition (see supplementary table S2). Classes did not differ with respect to their network characteristics (i.e. density, centrality, transitivity and reciprocity. See Supplementary Notes: Network Analysis). The mean valid wear time per day was 814 $CI_{95\%}$[633, 996] and this did not differ significantly between conditions (Kruskal-Wallis42, χ^2 (4)= 9.042, p_{value}=0.0601). Overall, participants wore accelerometers significantly longer (≥60 min) on weekdays than on weekend days (Mann-Whitney non-parametric ranking43

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

test (henceforth M-W): $z_{weekend}$=5.042, p_{value}<0.001). All tests used in this analysis (including the above two) are two-tailed.

Cut-points identified by Evenson and colleagues38 were used to classify time spent in sedentary (≤100 c.p.m.), light (101 - 2295 c.p.m.), moderate (2296-4011 c.p.m.), or vigorous intensity physical activity (≥4012 c.p.m.) (fig. 2). To check the robustness of our main findings, different cut-points were used to classify the same data (see supplementary fig. S2). We used minutes of moderate intensity and above (moderate-to-vigorous physical activity; MVPA) as a comparability measure36. Because of the sex-specific design and sex differences observed in MVPA (M-W for all and p_0 periods: z_{all}=7.238, z_{p0}=4.948; p_{values}<0.001), the data was analyzed and presented separately for boys and girls (fig. 3). Significant daily MVPA fluctuations of the actual mean MVPA (black line) (Kruskal-Wallis test42 for difference among different week days; $\chi^2(6)$=72.157, p_{value}<0.001) have been smoothed out by displaying the estimates of a locally (bandwidth: 0.3) weighted regression of MVPA on experimental days (red line). Fig. 3 shows the significant differences in children's MVPA on different days of the week, different weather conditions, and between holidays and school days (see M-W tests in supplementary material). These persistent environmental effects render statistical adjustment necessary.

In fig. 4 we averaged out all these effects by pooling data from both phases (A) and all experimental periods (B) for each of the different experimental conditions (by sex). To account for the different number of observations in the two experimental phases, we weighted each observation with the inverse of its probability of being sampled in experimental phase 1 or 2 (i.e. assuming the same number of observations by treatment and sex in each phase, see supplementary fig. S3 for non-balanced sample). The estimates of the locally (bandwidth: 0.3) weighted regression of MVPA on experimental days are presented in part A, with dotted left and right line-extensions indicating preliminary (p_0) and post-experimental (p_5) periods, respectively. Information from these two non-incentivized periods is not included in the calculation of the pooled average MVPA from all periods, illustrated by bar graphs (part B). Red error bars correspond to 95% confidence intervals of the respective subsamples (see also supplementary fig. S4 for box-plots). M-W tests on MVPA between different treatments (see supplementary table S3) confirm that for boys, team rewards consistently and significantly elicit higher levels of MVPA than the control scheme in both experimental phases. For girls, this was the case for the direct and indirect reciprocity schemes.

We used a multilevel hierarchical model to control for all observed (weather, weekends, holidays) and unobserved environmental effects in a more systematic way and allow the variance of the unit effects to be estimated conditional on the data and parameters at all different nested levels. In particular we estimated a 3-level mixed effects model in which repeated daily measurements (level 1) were nested within the same child (level 2) who was nested in one of the classrooms (level 3) participating in the study. Four different dummies, one for each different condition (IV-individual, DI-Direct Reciprocity, IND-Indirect reciprocity, TE-Team rewards) were included in the fixed part of the model together with the dummy variable *fem* indicating females and the continuous variable *val_time* indicating the daily minutes of valid accelerometry wear time. The control condition (random rewards) was omitted and functioned as a baseline level. Cross-classification of the repeated within-subject measurements was also handled in the fixed part of the model by including 73 day-dummies, Dum_t (i.e. ((37) valid experimental days * (2) experimental phases) -(1) the baseline day). Most importantly, using time-dummies in the fixed part of the model also served for controlling for all observed (weather, holidays, weekend etc.) or unobserved environmental effects.

With regard to the random part of the model, in the classroom level-3 we included a random coefficient for sex ($u_{0j}+u_{1j}*fem_{jt}$) accounting for the different sex composition of the different classrooms and the sex-specific weekly relative scores and thresholds. In the child level-2 we simply included a random intercept υ_{0ji}. Finally, we accounted for heteroskedasticity by classroom in the level-1 repeated measurements by adding 19 independent residual errors (ε_{jit}), one for each classroom (i.e. $\forall j: \varepsilon_{jit} \sim N(0, \sigma^2_{\varepsilon j} I)$), so that $E_{jit} = \sum_{j=1}^{19} \varepsilon_{jit}$) (see Methods and Supplementary methods). The resulting model, additionally including the interactions between the different conditions and sex, is described by the following equation:

$$\begin{aligned} MVPA_{jit} = {} & \beta_0 + \beta_1 IV_j + \beta_2 IND_j + \beta_3 DI_j + \beta_4 TE_j + \beta_5 fem_i + \beta_{105} IV_j fem_i + \beta_{205} DI_j fem_i \\ & + \beta_{305} IND_j fem_i + \beta_{405} TE_j fem_i + \beta_6 val\_time_{it} + \sum_{t=2}^{74} \beta_{t+5} Dum_t + u_{0j} + u_{1j} \\ & * fem_{jt} + \upsilon_{0ji} + E_{jit} \end{aligned} \quad (1)$$

Table 1 shows the coefficients, standard errors and 95% confidence intervals of the main variables (IV, DI, IND, TE, fem) of the fixed part in eq.1 (see supplementary text for the complete regression analysis). In M1, team incentives and direct reciprocity have a positive and significant effect on MVPA (β_4 = 4.93, p_{value}=0.001, $CI_{95\%}$:[1.99, 7.88] and β_3 = 4.30, p_{value}=0.010, $CI_{95\%}$:[1.05, 7.56] respectively), corresponding to a respective 67.1% (β_4/β_0) and 58.5% (β_3/β_0)) improvement as compared to the control condition (captured by β_0). Indirect reciprocity also had a positive effect, but this was marginally non-significant (β_2=2.69, p_{value}=0.061, $CI_{95\%}$:[-0.13, 5.50]). Pooling data from all social reward schemes together (IND, DI and TE) into one common regressor in M2 we found that in general social incentives had a positive and significant ($\beta_{2\text{-}4}$=3.81, p_{value}=0.001, $CI_{95\%}$:[1.50, 6.12) effect on MVPA, with a 51.8% higher value compared to the control condition. In contrast, MVPA in the IV scheme was found not to be significantly different (β_1=2.30, p_{value}= 0.093, $CI_{95\%}$:[-0.38, 4.97) from the MVPA in the control condition.

Including interaction terms between different conditions and sex (M3), we found that the effect of the team rewards on boys' MVPA was very high and significant (β_4=8.77, p_{value}<0.001, $CI_{95\%}$:[4.63, 12.92), corresponding to a 131.5% (β_4/β_0) increase compared to the control condition (β_0). While for girls the effect of team rewards was non-significant (β_4

$+ \beta_{405}$=2.53, $\chi^2(1)$=2.02, p_{value}=0.1555), the direct-reciprocity scheme affected girls' MVPA significantly ($\beta_3 + \beta_{305}$=5.441, $\chi^2(1)$=5.82, p_{value}=0.0159) and positively (76.5% improvement compared to control (($\beta_3 + \beta_{305}$)/($\beta_0 + \beta_5$))). Direct reciprocity did not affect boys' MVPA and indirect reciprocity was also non-significant for both boys and girls (only marginally ($\beta_2 + \beta_{205}$=3.195, $\chi^2(1)$=3.27, p_{value}=0.0706) for the latter). When combining all social rewards conditions in M4, we found that social reward schemes had a positive and significant effect both in boys' (β_{2-4}=4.375, p_{value}=0.002, $CI_{95\%}$:[1.60, 7.15]) and girls' ($\beta_{2-4} + \beta_{205-405}$= 3.346, $\chi^2(1)$=6.15, p_{value}=0.0132) MVPA. The individual incentives however did not significantly impact MVPA for either sex. The significant increases in MVPA in males in team scheme (β_4=11.20, p_{value}<0.001, $CI_{95\%}$:[6.23, 16.18) and in females in direct reciprocity scheme (β_3=4.21, p_{value}=0.040, $CI_{95\%}$:[0.20, 8.23]) are confirmed when running eq.1 regression model for the two separate sex subsamples (column M53 to M56 in supplementary table S5). Finally, testing for post-experimental effects (see supplementary methods and table S7), we found that while MVPA in all social incentives return to their initial levels (i.e. no significant differences between $MVPA_{pre}$ from p_0 and $MVPA_{post}$ from p_5), interestingly, the $MVPA_{post}$ in individual scheme was significantly lower than the corresponding $MVPA_{pre}$. As we discuss later this result is probably partly driven by students' enthusiastic and over-reactive behaviour at the beginning (p_0) of the study due to the Hawthorne44 or demand45 effects.

Discussion

Two main points may be inferred from our findings. First, unlike individual schemes, social reward schemes were found to have a significant overall effect in changing preadolescents' physical activity behaviour, translating into 24.4 additional minutes of MVPA per week (as compared to the control). Second, the different social reward schemes have asymmetric effects in boys and girls. Boys were more susceptible to the team incentives, achieving an additional 59.2 minutes of MVPA in a week compared to controls, while girls were more receptive to the direct reciprocity scheme, achieving an additional 36.9 minutes of MVPA per week. It is important to stress here that these two findings are net from any design-effects already introduced (but not measured) in the control condition (e.g. weekly point system and final rewards, performance feedback, relevant positioning in the classroom, game spirit, accelerometer use, etc.). Moreover, through regression analysis we statistically controlled for any environmental effects which could potentially influence participants (e.g. weather, holidays, etc.) or school-classes (e.g. influence by school directors, teachers, etc., also known in the literature as Hawthorne effects44).

Although we cannot claim that our design entirely excludes self-interest (incentives which are directly linked to individual payoffs) from the social rewards schemes (e.g. participants may comply to "peer-pressure" for increasing their individual payoffs by reducing "social" sanctions), the results from individual reward scheme indicate that self-interest itself is not a significant driver of behavioural change in this situation. In contrast, induced peer-pressure (likely together with other confounding factors, such as image scoring46, and observability power9) has an additional effect on the participants' time spent in MVPA as seen with the social rewards schemes. To understand the sex differences between the different social rewards schemes we should focus on the interplay between the different interactions induced (by design) and the different characteristics of the pre-existing sex-segregated networks.

Boys appear to maximize their MVPA performance when exposed to the team-competitive environment of the team rewards scheme, a result which is in agreement with a well-established finding in the literature[6,7] that male preadolescents are in general more interested than females in playing team games and forming larger networks[5, 47]. Interestingly, our within-classroom network analysis (see Supplementary Notes: Network Analysis) also shows that, compared to girls, boys are more frequently characterized as good or best friends ($\tau_{ij} \geq 4$) and have a more central role (higher closeness) both in the overall and in the within sex-specific classroom network.

In contrast, girls maximized their MVPA performance under the direct reciprocity scheme. This result can also be related to the preference of preadolescent females for small intimate groups[6, 4], usually including two or three reciprocal friendships[48]. In our dataset we also find that although girls have fewer "best-friendships", these friendships are more frequently reciprocated than in boys. We would argue that as direct bilateral interactions are more common within girls' intimate friendships, girls were therefore more responsive to direct reciprocity interaction schemes than boys.

As a robustness check we repeated our regression analysis (see supplementary table S6) using (a) an alternative activity-time classification, characterizing MVPA according to the cut points proposed by Freedson and colleagues[49, 50], (b) a less conservative filter including the accelerometer-delivery days (Wednesdays) with at least 5 hours of valid wear time (1359 additional observations) and (c) an additional control for the initial physical activity levels of the participants. In all cases, our sex-specific findings were retained.

Our study used randomly chosen independent small social networks to compare different incentive schemes and interaction structures for increasing physical activity in preadolescents; however, the work does have limitations. Although schools were located in different villages, they were all from a small geographical zone which makes it difficult to prevent contamination and also challenges the external validity of the study. On the other hand, recruiting subjects from the same area also reduces socio-economic differences between schools (at an aggregate level), and minimizes other environmental effects. At a disaggregate level though (students), we were unable to control for potential socioeconomic differences, which could be affecting the effectiveness of the network intervention. For instance, during the "pairing" process we took into account of the sex-based homophily but we were unable to take other characteristics into account (e.g. socio-economic status, racial or ethnical identification). This could have potentially affect the structure of the network (due to homophily) and therefore affect the intervention outcomes.

Moreover, the existence of other within-classroom environmental factors related to the methods, and most importantly with the conductors or monitors[44] of the experiment, should also be treated with caution as they might have an impact on children's behaviour, especially at the start of the study. However, we argue that these reactivity effects[51] decline over the course of our 7-week study and substituted by the intended effect of our interventions. To address this point specifically, we repeated the analyses using only the data collected during the third period of the experiment as subjects are already exposed to the monitors for three weeks, there is no end-of-the-game effect and it lasts for two consecutive weeks (with no feedback or researcher contact). Here, our main sex-specific results were also retained (see supplementary table S6 column M64). Finally, due to significant sex differences in MVPA, it was impossible for us to test the possible implications of our network interventions on mixed-sex groups, which may be a topic of interest for future research.

Our design also allows us to draw conclusions on post-experimental effects as we measured physical activity for an additional week after the intervention period (when subjects did not receive additional points or prizes). As in many other intervention studies focusing on post experimental effects51, 52 and on the effectiveness of intrinsic (vs. extrinsic) incentives16,17, we also found that physical activity levels in our social incentives schemes returned to pre-experimental levels (no significant differences between $MVPA_{pre}$ and $MVPA_{post}$, see column M72 of supplementary table S7). This result primarily suggests that our intervention did not create lasting change in norms but was completely contingent on the social rewards scheme. However, our post-experimental analysis reveals an additional information; activity levels of those in the individual incentive scheme were significantly lower in the post-experimental as compared to pre-experimental period (-7.57 minutes per day (p_{value}=0.002, $CI_{95\%}$[-12.37, -2.78]). In contrast, participants in the social incentives schemes simply returned to their initial physical activity levels. This implies that at least in relative terms, social incentives were more sustainable than individual incentives.

As the primary focus of this trial was not to study the persistence (or not) of post-experimental effects, we are unable to provide evidence-based explanations for these findings but instead hypothesize on the reasons. The researchers observed and teachers confirmed that in the pre-experimental period students were very enthusiastic in participating and particularly interested in and curious about their "magic box" (accelerometer). This observation gives reason to believe that pre-experimental performance might be higher than "normal" performance due to Hawthorne44 and demand effects45. This would imply that, compared to "normal" levels, the post-experimental performance might be actually higher in social incentives (and comparably, not that low in the individual incentives). As our design does not allow controlling for "normal" physical activity level, this is only a conjecture based on our observations in the field which would need to be tested with a different experimental design.

Our network interventions were conducted in two steps. In the first step we elicited within-classroom social networks and fixed the interaction structures (between same-sex friends of the same pairs or teams) which remained "locked" until the end of the study. In the second step we simply induced peer-pressure among friends within the locked-structures by attributing and exchanging points as a result of engaging in MVPA. It is likely that the interaction structures in the social rewards schemes would normally change after the first interactions if they were not enforced by-design. For instance, high performers would most probably seek to team up with other high performers in order to achieve higher rewards, or popular students could have benefited from their extensive network by grouping with high-performing best friends, exchanging in a way their social capital for a higher return in terms of game-points26–28, 15. By using locked interaction structures we sought to create a situation in which peer-pressure was the only reaction mechanism for any child seeking to increase his/her rewards; s/he simply had to encourage his/her partners to be more active. It is important to add that we did not observe any bullying effects against the weak performers. In contrast, teachers informally reported an increase in the quality of social behaviour among children.

The results indicate that sex-specific network interventions might be a powerful tool to positively change lifestyle behaviours of preadolescents, with boys being more susceptible to team interaction structures and girls to dyadic reciprocal structures. Changing usual behaviour and habits is complex, and this achievement may signal that this approach may have different applications in tasks of equal complexity. One needs to be aware however that our network interventions may not be that effective in eliciting change in behaviours of a different nature. For instance, in our study we took advantage of the inherited competitive nature of physical activities (especially in the school environment) which may not be relevant for other sorts of behavioural change (e.g. substance use). In a more abstract context, carefully designed network interventions based on the specific characteristics of the social networks and the nature of the bilateral relations and interactions may constitute a powerful tool for achieving behavioural change.

Methods

Recruitment

Students were recruited from 19 fifth-grade elementary-school classes in the same geographical region in Italy to minimize environmental effects. Over 98% of children provided signed parental informed consent (supplementary fig. S5) and participated in the study (Four parents did not sign the informed consent and they children did not participate in the study. One student dropped out due to health issues (unrelated to the study).

No prior sample size calculation was performed. The size of the sample was determined by the projects capacity and practical restrictions such us the number of available accelerometers (about 180 devises) and the number of recruited classes (19). To increase the sample size, the intervention was repeated twice (winter and spring experimental phases). Supplementary table S2 shows these sub-sample sizes by gender, classes and phases. The final sample size (and the respective size of sub-groups by gender, classes, experimental conditions and phase) was sufficient to perform non-parametric (Mann- Whitney, Wilcoxon and Kruskal-Wallis) and parametric (multilevel hierarchical model) analyses and testing, even when the data was collapsed at subjects level (one observation per individual).

Network Data Elicitation and Teams-Pairs formation

The strength of all dyadic ties (τ_{ij}) between classmates i and j was elicited through a teacher-administered self-reported questionnaire in which participants rated individually each one of their classmates (see supplementary fig. S6) from 1 "… is not my friend" to 5 "…is one of my best friends" ($1 \leq \tau_{ij} \leq 5$). In the supplementary fig. S7 we use modern Multidimensional Scaling54 to visualize spatial distance only between best friends ($\tau_{ij}=5$). Separate colours highlight boys and girls, while we also illustrate all friendships of $\tau_{ij}\geq3$. These graphs demonstrate that most networks are sex-segregated.

The network data was used for the formation of interacting teams or pairs only in the social reward schemes. Only same-sex classmates with relatively strong ties ($\tau_{ij}\geq\tau_{ji}\geq3$) were randomly matched together (only two exceptions where $\tau_{ij}\geq3\geq\tau_{ji}$). Considering that school networks were characterized by sex homophily (see supplementary fig. S7-S8) and that students would probably self-select to pair or team up with same-sex colleagues, we followed this natural predisposition and induced same-sex matching that also allowed us to control for known sex differences on children's physical activity levels36. Matching relatively strong-tied classmates was purposely designed to maximize peer-pressure between relatively close friends24 and also to minimize the possibility of any adverse effects between distant friends. Pseudo-randomisation was preferred to self-selection matching to ensure compatibility in the matching process for all classmates even in the most complicated interaction schemes (i.e. indirect reciprocity).

Experimental Phases, Periods and Conditions

The study was conducted in two experimental phases in 2014, with 10 classes (from 7 schools, classes from the same schools were participating in the same condition) participating in the winter phase (N_w=176) and 9 different classes (from 8 schools) in the spring phase (N_s=173). Principal teachers served as contact points between students and researchers throughout the whole experiment to preserve students' anonymity (teachers were not informed about the performance of the students throughout the whole experiment, see supplementary fig. S9 for teachers' instructions). In both phases, after a 1-week preliminary period (p_0), schools were randomly assigned to different intervention conditions which were conducted in parallel for 5 more consecutive weeks divided into 4 periods: three 1-week (p_1,p_2 and p_4) and one 2-week (p_3) period. As researchers required knowledge about the experimental condition in order to prepare and deliver the corresponding feedback, data collection and analysis were not performed blind to the conditions of the experiments. However, the main outcome measure is objectively measured and therefore researcher bias will have unlikely to have influenced the outcome assessment. The study concluded with a final 1-week post-experimental period (p_5) in which students did not receive any kind of incentive (points). Classification medals were given virtually and corresponding to 100, 80 and 60 points in all but the control condition (in which the random allocation of points was used). The different experimental conditions implemented in the classes defined the points allocation rules among classmates. In the individual reward scheme (IV) participants received the points corresponding to their own classification medal. In the social rewards conditions, though, points were not allocated to the "self" but to other classmates (fig. 1), who were bilaterally satisfying (with two exceptions) the condition $\tau_{ij}\geq\tau_{ji}\geq3$ (for facilitating analysis henceforth we call them as "partners"). In the direct (DI) and indirect reciprocity (IND) schemes each individual was engaged in two active and two passive interactions; s/he a) was sending half of her/his points to one partner (point-taker) and b) the remaining half to another point-taker while s/he c) was receiving the half of the points of one partner (point-giver) and d) the half of the points of another point-giver. In the IND each participant was interacting with four different partners while in the DI they only interacted with two partners who were at the same time both point-givers and point-takers. Open network structure was imposed on all same-sex networks assigned to DI, by preventing an individual's point-givers/takers to interact with each other. However, in the IND, the open network structure condition was satisfied only for those same-sex IND networks with N≥10 (4 out 6). For the remaining two IND networks, we allowed for interaction between point-givers and point-takers but never within point-givers or within point-takers. Finally in the team scheme (TE), three partners were assigned to the same team and engaged in a Public Goods Game39 which the points of all team members were gathered together and then shared equally among them.

Instructions of the Game (accelerometers and conditions)

Special attention has been given to the explanation of the instructions both to the teachers and to the students. Teachers received general instructions about their role in the conduct of the experiment during the recruitment period. The supplementary fig. S9 shows the written instructions given to the responsible teachers. Prior to the experiment, researchers also organized an interview with each one of the teachers in order to be sure that they accurately understood the general instructions. Moreover, researchers met teachers in the school on a weekly base during the collection/delivery of the accelerometers and answered any questions raised by them or the students.

For the instructions to the students we used a "game" language (e.g. we use the term "Magic Box" instead of accelerometer) which was easier to be understood by students. Instructions were given in two different stages in order for them not to be overwhelmed with new information: A) Participants were instructed at the beginning of p_0 to wear the accelerometers on a belt on their right hip, during all waking hours (except when performing water or contact sports) throughout all periods (and p_5), including weekends. Additional to the written illustrated instructions (see supplementary fig. S10), researchers gave a small demonstration in the classroom for the correct use of the accelerometers. B) Incentive-instructions and classmates grouping/matching (where applicable) were revealed to students at the beginning of p_1. In the social incentive schemes students received their partners' names and were described in detail how their behaviour can potentially affect their partners' points. In addition, we attached on each child's accelerometer a personalized key ring which was indicating the names of his/her partners (when applicable). In supplementary fig. S11 we show the exact information received by the students.

At the beginning of period p_2 when researchers (through teachers) gave the students their first feedback (corresponding to their performance in p_1), there were only few cases in the social incentive schemes where students reported that they had not understood the interaction scheme with their partners. However, the further information given at that stage (see supplementary fig. S13) was enough to erase any confusion, as in the next week (beginning of period p_3) teachers did not receive any further questions regarding the interaction schemes and assured the researchers that all students already knew their partners and were aware of how the points were distributed or exchanged. The students participating in the random rewards scheme found it most difficult to understand that their points were independent of their performance. However, after receiving further information at the beginning of p_1, this was resolved. Supplementary table S6 (column M64) demonstrates that our main results were retained even after excluding these potentially "noisy" periods.

Physical Activity Measurement and Feedback

Physical activity was measured by Actigraph accelerometers (either GT1M, GT3X or GT3X+; ActiGraph, Pensacola, FL.) and was communicated individually to classmates as minutes of intensive physical activity (examples were given in the classroom and from p_1 participants could also observe it in the provided histogram-feedback). Accelerometers were handed out at the beginning of each period (Wednesday, 12:00), and collected at the end of the period (Tuesday, 08.00) for data extraction, recharging and feedback preparation (see supplementary fig. S1).

At the beginning of each period (for all $p_{t\geq 1}$) participants received individual information (see supplementary fig. S12-S13) including:

a) a histogram of their average hourly MVPA minutes in p_{t-1} and their average individual period-score ($s_{i,pt-1}$),

b) the upper (T_t^H) and the lower $\left(T_t^L\right)$ within-classroom thresholds defined for the current period p_t for boys and girls separately and coinciding to the upper ($Pr[S_{t-1}>s_{t-1}]\leq 2/3$ and $Pr[S_{t-1}\leq s_{t-1}]\geq 2/3$) and lower ($Pr[S_{t-1}>s_{t-1}]\leq 1/3$ and $Pr[S_{t-1}\leq s_{t-1}]\geq 1/3$) tertiles of boys and girls score-distributions (S_{t-1}) in p_{t-1}.

c) Their relative classification in p_{t-1} falling into one of the following three categories: Golden if $s_{i,t-1} \geq T_{t-1}^H$, Silver if $T_{t-1}^H > s_{i,t-1} \geq T_{t-1}^L$ and Bronze medal if $s_{i,t-1} < T_{t-1}^L$.

d) The points obtained based on their (or on their partners') performance in p_{t-1}. In the social reward schemes, participants were also informed about the medals and points of their partners. This 3-category medal system was designed in order to dampen competition (after p_1 all students could potentially reach T_t^H and win the gold medal)) which, if too extreme, could cause negative side effects (e.g. hostile behaviour towards high performers). This reward system proved successful as teachers reported no conflicts among students but an increase in the team spirit especially for the ones participating in the team condition.

Instruction intelligibility among 9-11 year old students was a key challenge of this study due to the complexity of the interaction schemes. However, we managed to address this issue by a repeated interaction with teachers and students before and during the experiment and by offering clear instructions and feedback on a regular basis. In supplementary fig. S1 we show the timeline of the experiment, highlighting in red letters all the occasions when researchers interact with teachers and students during the experiment. After p_1 teachers reported that students did not have any more questions and they fully understood the rules of the game.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Experimental Payments

Additional to the points earned in p_1-p_4, participants received a 100-points bonus at the end of p_0 for correctly wearing accelerometers. Irrespective of condition, points from all periods ($p_{t\geq0}$)were summed at the end of the experiment (during p_5) for each participant individually who could exchange them with one or more items from a gift-list including 94 items of different points-value (see supplemetary fig. S14). This was also true for the control condition, in which students were earning 100, 80 or 60 points with 1/3 wining probability after a lottery draw at the end of each experimental period (p_1-p_4). The expected payoff per period (for all $p_{t\geq1}$) for a random student in all conditions, including the control, was equal to 80 points, with the total expected payoff from all periods reaching ($p_{t\geq0}$) 420 points (min=340 points, max=500 points). By providing all participants with the same-size material incentives we ensured that any resulting differences between students participating in different experimental conditions would be due to the different intervention scheme and not to the size or the material incentive itself. Hence, material incentives were included as a methodological tool for quantifying physical activity performance while their effectiveness55 does not constitute an objective of this study.

Physical Activity Data Collection and Filtering

Accelerometers were set to record at 10-second epochs, and data from the accelerometers were downloaded using ActiLife software (version 5.7.4, full edition). Data before 07:00 and after 23:00 were excluded and non-wear time was removed, using a criterion of consecutive runs of zero counts per minute (cpm) for a minimum duration of 60 minutes, allowing for a ‘spike tolerance’ of 236, 38. Remaining data were included if accelerometer wear time was at least 10 hours per day. The application of this last filter basically removed all delivery and recollection days and left our sample with a total of 9776 valid days (see supplementary fig. S15-S16 and table S1). At least three independent school-classes were assigned to each experimental condition with an average of 35.4 females and 34.4 males participating in each condition (see suplemtary table S2). The mean number of valid days per individual was 28.1 (SD) with team scheme (TE) only being significantly different (p_{value}<0.01 in all between-conditions bilateral Mann-Whitney non-parametric ranking43 tests including the TE) to the other schemes mainly due to the late involvement of two TE classes. The mean number of valid minutes per day was 814 (SD), which did not differ significantly between conditions (Kruskal-Wallis χ^2= 9.042 with 4 d.f. and p_{value}=0.0601). However, accelerometers were worn for significantly longer (≥60 min) on weekdays as compared to weekend days (M-W: $z_{weekend}$=5.042, p_{value}<0.001).

Regression Analysis

We performed regression analysis to control for any environmental effects in a systematic way. The dependent variable was the total daily minutes in MVPA. In order to capture treatment effects we used the following four dummy variables as the main regressors for our model:

- *Individual (IV)*: takes value 1 for all those measurements recorded during the implementation of the individual rewards scheme, otherwise 0.

Europe PMC Funders Author Manuscripts

- *Direct (DI)*: takes value 1 for all those measurements recorded during the implementation of the direct reciprocity scheme, otherwise 0.
- *Indirect (IND)*: takes value 1 for all those measurements recorded during the implementation of the indirect reciprocity scheme, otherwise 0.
- *Team (TE)*: takes value 1 for all those measurements recorded during the implementation of the team rewards scheme, otherwise 0.

Note that these dummies take the value 0 for all those measurements recorded during p_0 and p_5 when all children either received the 100 points bonus or not received any points at all. The dummy variable *control* taking the value 1 for all those measurements recorded during the implementation of the control scheme is used as control group for our regressions and therefore omitted from the regression table.

We used a multilevel model (using maximum-likelihood estimator) to allow the variance of the unit effects to be estimated conditional on the data and parameters at all different nested levels (school-class(j) > student(i) > measurements(t)). In particular we used a 3-level mixed effects model in which repeated daily measurements (level 1 t-level) are nested within the same child (level 2 or i-level), who is nested in one of the classes (level 3 or j-level) participating in the study. Treatments' dummies were included in the fixed part of the model together with the dummy variable *fem* indicating females and the continuous variable *val_time* reporting the number of validated daily minutes. Note that we filtered out all data with *val_time≤600 minutes*. Cross-classification of the repeated within-subject measurements was also handled in the fixed part of the model by including 73 day-dummies, Dum_t (i.e. ((37) valid experimental days * (2) experimental phases) -(1) the baseline-reference day). Using time-dummies (Dum_τ, $\tau \in [1,74]$) in the fixed part of the model also enabled controlling for all observed (weather, holidays, weekend etc.) or unobserved environmental effects.

With regards to the random part of the model, at the class-level we included a random coefficient for sex (fem) to account for the differences in sex composition of the different classrooms, expecting classes with a high percentage of males to have a higher average classroom performance. The random sex coefficient also accounted for the fact that although boys and girls in the same classroom were allocated to the same condition, their relative performance and weekly thresholds were sex-specific. We also allowed for the random effects at level-3 to be correlated (general symmetric covariance matrix). At level-2, random effects accounted for heteroskedasticity by classroom at the measurements-level (see supplementary table S4, column M43). Note that heteroskedastic residuals at this level also captured phase-heteroskedasticity since each class participated in either the first or in the second experimental phase. See supplementary material for regression model and post-hoc testing and diagnostic. Data distribution in this analysis was assumed to be normal but this was not formally tested. In supplementary figure S18 we show the actual residuals distribution. As compared to the normal distribution, the distribution of our residuals is slightly skewed on the left (0.83) and leptokurtic (5.56).

Europe PMC Funders Author Manuscripts

Ethical Approval

Ethical approval was given by the Experimental Economics Research Ethics Committee, Joint Research Centre, European Commission. The regional representative office of the Italian Ministry of Education (Ufficio Scolastico Territoriale per la Lombardia) authorised the study in the schools of its territory and coordinated the collection of parental informed consent. The experimental design was also approved by the European Data Protection Supervisor who approved the study in respect to the data protection rights of the participants. All relevant approvals and documents are available upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank S. Gächter, A.M. Espin, S. Kriemler, F. Exadaktylos, E. Yoeli, E. Woelbert, S. Stoffel and J. Fooken for invaluable discussions and feedback. We also thank Dr. Claudio Merletti representing the Ufficio Scolastico Territoriale per la Lombardia for his authorization and ethical approval for the realization of this study in schools under his responsibility. We are especially grateful to Mrs. Paula Benetti from the same office for coordinating the communication with all school directors, teachers, parental associations and students' parents. We finally thank all the school directors and teachers involved in the study and all the parents and especially the children who agreed voluntarily to participate in this study (Scuola Primaria "G. Galilei", Ispra; Scuola Primaria "S. Pellico", Ranco; Scuola Primaria "G. Pascoli", Taino; Scuola Primaria "D. Alighieri", Angera; Scuole Primarie "G. Ungaretti" and "G. Matteotti", Sesto Calende; Scuola Primaria "D. Alighieri" Golasecca; Scuola Primaria "A. Manzoni" Mercallo; Scuola Primaria "L. Scotti" Laveno; Scuola Primaria" G. de Amicis" Vergiate; Scuola Primaria "G. Wojtyla" Cimbro; Scuola Primaria "S. Tamborini" Varano Borghi; Scuola Primaria "A. Liborio" Comabbio, Scuola Primaria "A. Manzoni" Malgesso). This research was exclusively funded by European Commission. The supply of the technical equipment (accelerometers) and the contributions of HB and EvS were supported by the Medical Research Council (MC_UU_12015/7), and the Centre for Diet and Activity Research (CEDAR), a UKCRC Public Health Research Centre of Excellence (RES-590-28-0002). Funding from the British Heart Foundation, Department of Health, Economic and Social Research Council, Medical Research Council, and the Wellcome Trust, under the auspices of the UK Clinical Research Collaboration is gratefully acknowledged. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Any opinions expressed in this article are those of the authors and not of the European Commission or any other involved institute affiliated by the authors. Data on physical activity and bilateral friendships are provided by the authors upon request.

References

1. Sinan A, Walker D. Identifying influential and susceptible members of social networks. Science. 2012; 337(6092):337–341. [PubMed: 22722253]
2. Centola D, Macy M. Complex contagions and the weakness of long ties. American Journal of Sociology. 2007; 113(3):702–734.
3. Valente TW. Network interventions. Science. 2012; 337(6090):49–53. [PubMed: 22767921]
4. Eckles D, Karrer B, Ugander J. Design and analysis of experiments in networks: Reducing bias from interference. Journal of Causal Inference. 2017; 5(1)
5. Benenson J, Apostoleris N, Parnass J. The organization of children's same-sex peer relationships. New Directions for Child and Adolescent Development. 1998; 1998(81):5–23.
6. Mehta MC, Strough JN. Sex segregation in friendships and normative contexts across the life span. Developmental Review. 2009; 29(3):201–220.
7. Rose AJ, Rudolph KD. A review of sex differences in peer relationship processes: potential trade-offs for the emotional and behavioral development of girls and boys. Psychological bulletin. 2006; 132(1):98. [PubMed: 16435959]
8. Carrell SE, Hoekstra M, West JE. Is poor fitness contagious?: Evidence from randomly assigned friends. Journal of Public Economics. 2011; 95(7):657–663.

9. Cohen-Cole E, Fletcher JM. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. Journal of health economics. 2008; 27(5):1382–1387. [PubMed: 18571258]

10. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. New England journal of medicine. 2007; 357(4):370–379. [PubMed: 17652652]

11. De La Haye K, Robins G, Mohr P, Wilson C. How physical activity shapes, and is shaped by, adolescent friendships. Social science & medicine. 2011; 73(5):719–28. [PubMed: 21802807]

12. Cialdini RB, Kallgren CA, Reno RR. A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. Advances in experimental social psychology. 1991; 24:201–234.

13. Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. Psychological science. 2007; 18(5):429–434. [PubMed: 17576283]

14. Morris S. Contagion. The Review of Economic Studies. 2000; 67(1):57–78.

15. Calvó-Armengol A, Jackson MO. Peer pressure. Journal of the European Economic Association. 2010; 8(1):62–89.

16. Deci EL. Effects of externally mediated rewards on intrinsic motivation. Journal of personality and Social Psychology. 1971; 18(1):105.

17. Gneezy U, Meier S, Rey-Biel P. When and why incentives (don't) work to modify behavior. Journal of Economic Perspectives. 2011; 25(4):191–210.

18. Valente, TW. Social networks and health: Models, methods, and applications. Oxford University Press; 2010.

19. Pickard G, Pan W, Rahwan I, Cebrian M, Crane R, Madan A, Pentland A. Time-critical social mobilization. Science. 2011; 334(6055):509–512. [PubMed: 22034432]

20. Centola D. An experimental study of homophily in the adoption of health behavior. Science. 2011; 334(6060):1269–1272. [PubMed: 22144624]

21. Leider S, Möbius M, Rosenblat T, Do AQ. Directed altruism and enforced reciprocity in social networks. The Quarterly Journal of Economics. 2009; 124(4):1815–1851.

22. Rand D, Dreber A, Ellingsen T, Fudenberg D, Nowak MA. Positive interactions promote public cooperation. Science. 2009; 325(5945):1272–1275. [PubMed: 19729661]

23. Nowak MA. Five rules for the evolution of cooperation. Science. 2006; 314(5805):1560–1563. [PubMed: 17158317]

24. Mani A, Rahwan I, Pentland A. Inducing peer pressure to promote cooperation. Scientific Reports. 2013; 3

25. Aharony N, Pan W, Ip C, Khayal I, Pentland A. Social fMRI: Investigating and shaping social mechanisms in the real world. Pervasive and Mobile Computing. 2011; 7(6):643–659.

26. Bandiera O, Barankay I, Rasul I. Team incentives: evidence from a firm level experiment. Journal of the European Economic Association. 2013; 11(5):1079–1114.

27. Nowak MA, Tarnita CE, Antal T. Evolutionary dynamics in structured populations. Philosophical Transactions of the Royal Society B: Biological Sciences. 2010; 365(1537):19–30.

28. Rand DG, Arbesman S, Christakis NA. Dynamic social networks promote cooperation in experiments with humans. Proceedings of the National Academy of Sciences. 2011; 108(48): 19193–19198.

29. Mountjoy M, Andersen LB, Armstrong N, Biddle S, Boreham C, Bedenbeck HPB, Ekelund U, Engebretsen L, Hardman K, Hills A, Kahlmeier S. International Olympic Committee consensus statement on the health and fitness of young people through physical activity and sport. British Journal of Sports Medicine. 2011; 45(11):839–848. [PubMed: 21836168]

30. Metcalf B, William H, Wilkin T. Effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively measured outcomes. (EarlyBird 54). 2012:e5888.

31. Babcock P, Bedard K, Charness G, Hartman J, Royer H. Letting down the team? Social effects of team incentives. Journal of the European Economic Association. 2015; 13(5):841–870.

32. Yoeli E, Hoffman M, Rand DG, Nowak MA. Powering up with indirect reciprocity in a large-scale field experiment. Proceedings of the National Academy of Sciences. 2013; 110(Supplement 2): 10424–10429.
33. Greiner B, Levati MV. Indirect reciprocity in cyclical networks: An experimental study. Journal of Economic Psychology. 2005; 26(5):711–731.
34. Charness G, Gneezy U. Incentives to exercise. Econometrica. 2009; 77(3):909–931.
35. Acland D, Levy MR. Naiveté, projection bias, and habit formation in gym attendance. Management Science. 2015; 61(1):146–160.
36. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Medicine & Science in Sports & Exercise. 2007; 40(1): 181–188.
37. Krackhardt D, Stern RN. Informal networks and organizational crises: An experimental simulation. Social Psychology Quarterly. 1988:123–140.
38. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. Journal of Sport Science. 2008; 26(14):1557–1565.
39. Ledyard, JO. Public Goods: A Survey of Experimental ResearchHandbook of Experimental Economics. Kagel, John; Roth, Alvin, editors. Princeton: Princeton University Press; 1995. 111–194.
40. Winkler EAH, Gardiner PA, Clark BK, Matthews CE, Owen N, Healy GN. Identifying sedentary time using automated estimates of accelerometer wear time. British Journal of Sports Medicine. 2012; 46(6):436–442. [PubMed: 21504965]
41. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Medicine & Science in Sports & Exercise. 2011; (43):357–364. [PubMed: 20581716]
42. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. Journal of the American statistical Association. 1952; 47(260):583–621.
43. Mann, HB, Whitney, DR. The annals of mathematical statistics. 1947.
44. French, J. Experiments in field settingsResearch Methods in the Behavioral Sciences. 1953.
45. Zizzo DJ. Experimenter demand effects in economic experiments. Experimental Economics. 2010; 13(1):75–98.
46. Wedekind C, Milinski M. Cooperation through image scoring in humans. Science. 2000; 288(5467):850–852. [PubMed: 10797005]
47. Baines E, Blatchford P. Sex differences in the structure and stability of children's playground social networks and their overlap with friendship relations. British Journal of Developmental Psychology. 2009; 27(3):743–760. [PubMed: 19994579]
48. Howes C, Chamberlain B, Lee L. Ethnic heterogeneity of social networks and cross-ethnic friendships of elementary school boys and girls. Merrill-Palmer Quarterly. 2007; 53(3):325–346.
49. Freedson PS, Pober D, Janz KF. Calibration of accelerometer output for children. Medicine & Science in Sports & Exercise. 2005; 37(11 suppl):S523–30. [PubMed: 16294115]
50. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. Medicine & Science in Sports & Exercise. 2011; (43):1360–1368. [PubMed: 21131873]
51. Dössegger A, Ruch N, Jimmy G, Braun-Fahrländer C, Mäder U, Hänggi J, Hofmann H, Puder JJ, Kriemler S, Bringolf-Isler B. Reactivity to accelerometer measurement of children and adolescents. Medicine and science in sports and exercise. 2014; 46(6):1140. [PubMed: 24219978]
52. Frey E, Rogers T. Persistence: How treatment effects persist after interventions stop. Policy Insights from the Behavioral and Brain Sciences. 2014; 1(1):172–179.
53. van der Linden S. The nature of viral altruism and how to make it stick. Nature Human Behaviour. 2017; 1(41)
54. Torgerson WS. Multidimensional scaling: I. Theory and method. Psychometrika. 1952; (17):401–419.
55. Camerer C, Hogarth RM. The effects of financial incentives in experiments: A review and capital-labor-production framework. Journal of Risk and Uncertainty. 1999; 19.1(3):7–42.

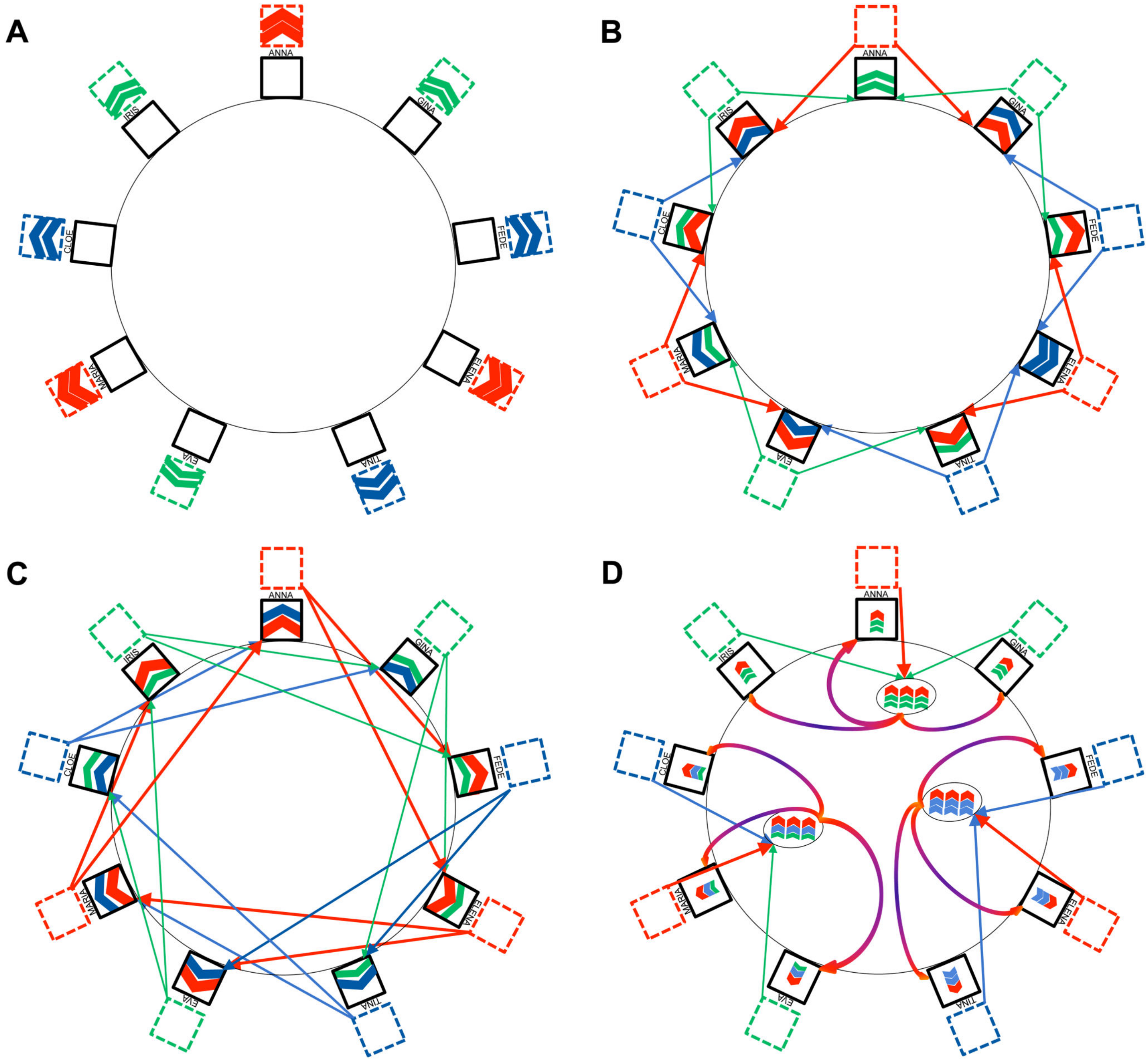

Fig. 1. Rewarding, interactions and allocation of points in the social-rewards schemes.

A hypothetical scenario of 9 classmates is presented. A) **Rewarding for physical activity performance.** Red, blue and green upper rectangles represent children's relative physical activity performance in the classroom in one period and correspond to 100, 80 and 60 points respectively. B-D) **Points allocation in Social-Rewards Schemes;** each red, blue and green slice (arrow) corresponds to 50, 40 and 30 points allocation (transfer) respectively. B) **Direct reciprocity scheme**; each child divides his/her points between two partners and receives half of the points that each one of these partners generated. C) **Indirect reciprocity scheme**; each child divides his/her points between two partners and receives half of the points that two other partners generated. D) **Team scheme**; the points of 3 partners belonging to the same team are added to a common pot and then shared equally among the partners.

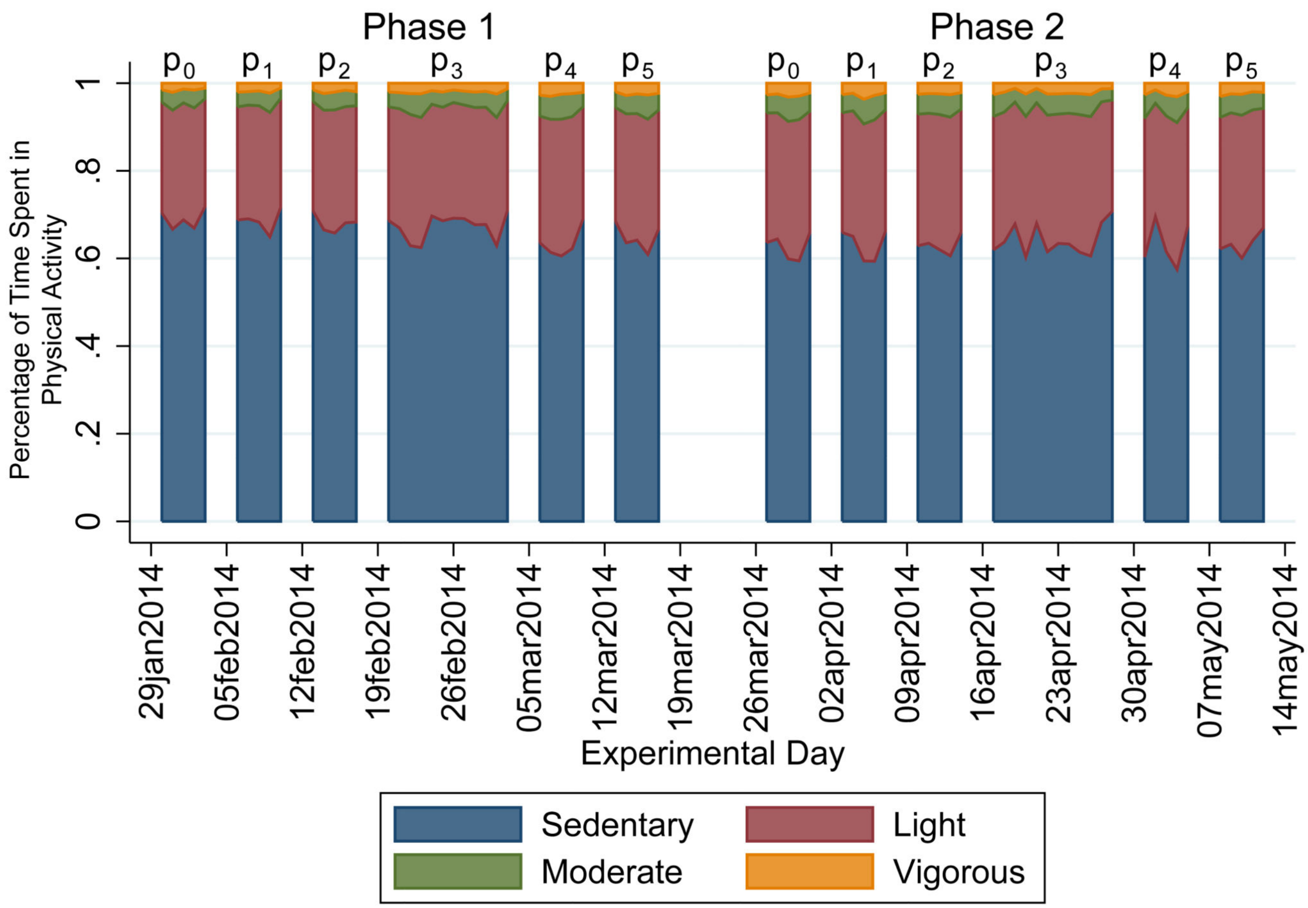

Fig. 2. Classification of Daily Activity-Time for all experimental conditions (n_{phase1} =176, n_{phase2}=173).

The classification uses Evenson and colleagues cut-points (30).

Europe PMC Funders Author Manuscripts

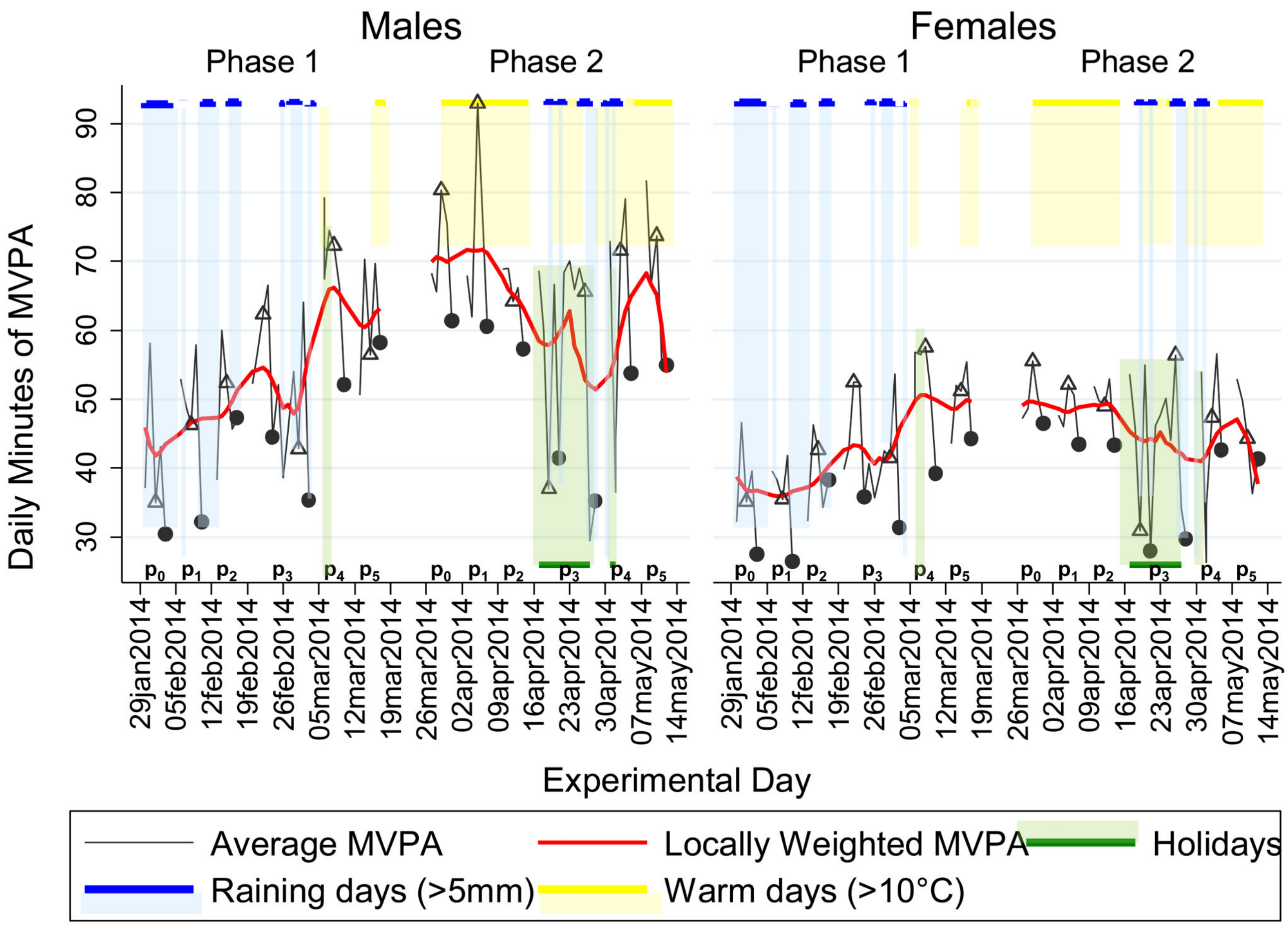

Fig. 3. Average Daily Minutes of Moderate to Vigorous Physical Activity (MVPA).

Data is aggregated and averaged at day-level for 92 and 85 females and for 84 and 88 males in Phases 1 and 2 respectively. Black line corresponds to the means of the actual MVPA while the red line are the estimates from a locally (with 0.3 bandwidth) weighted regression of MVPA on experimental days. Black dots and triangles indicate Mondays and Saturdays respectively. Blue, yellow and green shadowed days-periods highlight, rainy (precipitation more than 5mm), warm (temperature higher than 10°C) and holidays days/periods respectively.

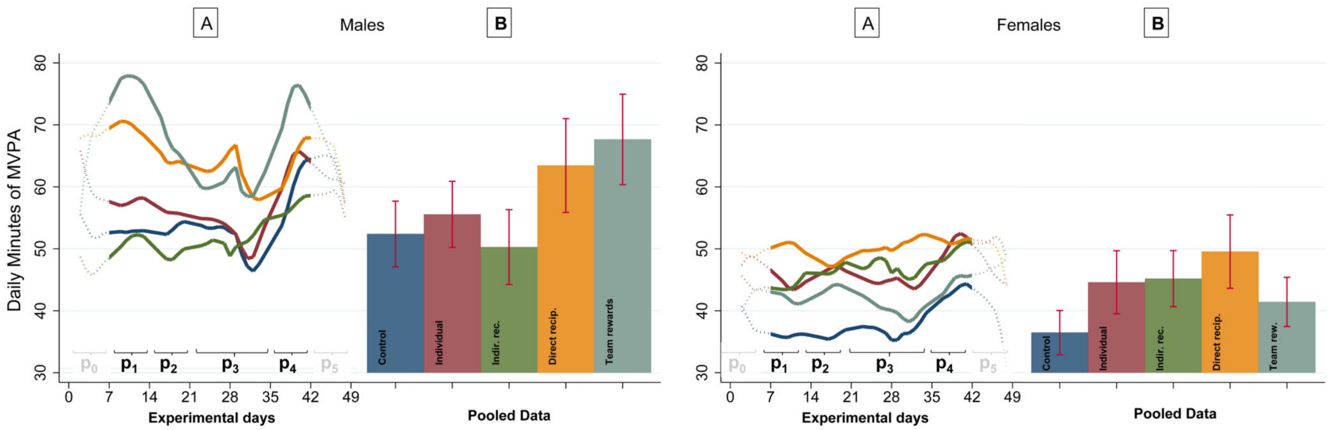

Fig. 4. Average Daily Minutes Moderate to Vigorous Physical Activity (MVPA) by experimental condition and sex (167 females and 162 males).

(A) Locally weighted estimations of MVPA on experimental days (pooled data from both experimental phases). Dotted lines correspond to non-incentivized periods. (B) Pooled data from all experimental periods (p_1-p_5) with 95% confidence interval error bars. (A-B) Each observation has been weighted by the inverse of its probability of being sampled in experimental phase 1 or 2.

Table 1

Multilevel Regression Models on MVPA

			(M1)	(M2)	(M3)	(M4)
FIXED EFFECTS Coefficients	Conditions	β_1 (IV)	2.32 (1.36) [-0.34, 4.98]	2.30 (1.37) [-0.38, 4.97]	1.85 (1.66) [-1.41, 5.11]	1.80 (1.68) [-1.49, 5.09]
		β_2 (IND)	2.69 (1.43) [-0.13, 5.50]	3.81** (1.18) [1.50, 6.12]	2.12 (1.87) [-1.54, 5.77]	4.37** (1.42) [1.60, 7.15]
		β_3 (DI)	4.30** (1.66) [1.05, 7.56]		3.47 (2.10) [-0.66, 7.59]	
		β_4 (TE)	4.93** (1.50) [1.99, 7.88]		8.77*** (2.11) [4.63, 12.92]	
		β_5 (fem)	-14.27*** (2.47) [-19.12, -9.43]	-14.34*** (2.47) [-19.19, -9.50]	-13.79*** (2.38) [-18.46, -9.12]	-14.11*** (2.54) [-19.10, -9.13]
	Interactions	β_{105} (IV*fem)			1.03 (2.01) [-2.92, 4.98]	1.03 (2.02) [-2.92, 5.00]
		β_{205} (IND*fem)			1.08 (2.23) [-3.29, 5.45]	-1.029 (1.448) [-3.87, 1.181]
		β_{305}(DI*fem)			1.97 (2.83) [-3.57, 7.52]	
		β_{405} (TE*fem)			-6.24* (2.46) [-11.06, -1.43]	
		β_0 (Constant)	7.01 (3.58) [0.004, 14.02]	7.35* (3.62) [0.26, 14.44]	6.67 (3.54) [-0.27, 13.61]	7.24* (3.63) [-0.14, 14.35]
Model Stats	Observations		9776	9776	9776	9776
	Wald χ^2		2063.2	2057.4	2078.1	2059.0
	$p_{value<}$		0.0001	0.0001	0.0001	0.0001

Notes: Minutes of Moderate to Vigorous Physical Activity (MVPA) is the dependent variable in all models. Standard errors and 95% confidence interval in parentheses and brackets respectively. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Four different dummies, one for each different condition (IV-individual, DI-Direct Reciprocity, IND-Indirect reciprocity, TE-Team rewards) and their interaction with sex (in M3 and M4) are included in the fixed part of the model. Control (valid time, date) variables, random effects parameters and residuals as described in eq.1, are omitted from the table for facilitating illustration. In M2 and M4, variables IND, DI, TE are collapsed to one variable (social rewards).