Acta Psychologica 215 (2021) 103291

Contents lists available at ScienceDirect

Acta Psychologica

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

Coordination effort in joint action is reflected in pupil size

Basil Wahn [a,b,*], Veera Ruuskanen [c], Alan Kingstone [a], Sebastiaan Mathôt [c]

[a] Department of Psychology, University of British Columbia, Vancouver, Canada
[b] Department of Psychology, Leibniz Universität Hannover, Hannover, Germany
[c] Department of Psychology, University of Groningen, Groningen, Netherlands

ARTICLE INFO

Keywords:
Pupillometry
Joint action
Multiple object tracking
Collaboration
Social cognition
Human-robot interaction

ABSTRACT

Humans often perform visual tasks together, and when doing so, they tend to devise division of labor strategies to share the load. Implementing such strategies, however, is effortful as co-actors need to coordinate their actions. We tested if pupil size – a physiological correlate of mental effort – can detect such a coordination effort in a multiple object tracking task (MOT). Participants performed the MOT task jointly with a computer partner and either devised a division of labor strategy (main experiment) or the labor division was already pre-determined (control experiment). We observed that pupil sizes increase relative to performing the MOT task alone in the main experiment while this is not the case in the control experiment. These findings suggest that pupil size can detect a rise in coordination effort, extending the view that pupil size indexes mental effort across a wide range of cognitively demanding tasks.

1. Introduction

In everyday life, humans frequently perform tasks together (e.g., when jointly shopping groceries or solving puzzles). In these tasks, humans often tend to distribute the labor. For instance, when shopping in the supermarket, one person may look for vegetables while another person may look for fruits. While such labor divisions clearly help performance efficiency, they also demand a coordination effort as co-actors need to coordinate their actions (e.g., by anticipating the partner's actions and accordingly adapting their own actions).

In recent years, several studies have investigated factors that influence labor division strategies in visuospatial tasks such as visual search (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; A.A. Brennan & Enns, 2015a, 2015b; Szymanski et al., 2017; Wahn, Czeszumski, Labusch, Kingstone, & König, 2020) or a multiple object tracking (MOT) task (Wahn & Kingstone, 2020; Wahn, Kingstone, & König, 2017; Wahn, König, & Kingstone, 2020) – see Wahn, Kingstone, and König (2018), for a recent review. For a MOT task, Wahn et al. (2017) investigated how the formation of division of labor strategies depends on the information that co-actors receive about each other. In particular, a pair of participants first view several objects presented on separate screens (one for each participant). A subset of these objects would turn gray to indicate that these are the target objects. Then these targets revert back to their original color and all the objects begin to move randomly across the screen for several seconds while participants attempt to track the targets. After the objects stop moving, each person selects the objects that they believe are the targets. The members then see the objects that were selected by their partner. Over the course of a few trials, participants typically devise labor division strategies with the most common strategy being a left and right labor division. That is, one person tracks the leftmost targets and the other the rightmost targets. Such division of labor strategies are beneficial because they allow pairs of participants to track more targets together than either of the pair members would be able to do alone; in other words, there is a clear group benefit (Wahn et al., 2017).

We have since replicated this left and right labor division preference in follow-up studies involving human-human (Wahn & Kingstone, 2020; Wahn, König, et al., 2020) and also human-computer pairs (Wahn & Kingstone, 2021). That is, for human-computer pairs, we found that participants will coordinate labor division also with a computer partner if the computer is initially described as behaving in a human-like manner, and if the computer adopts a human-like strategy in the MOT task (i.e., consistently tracking the leftmost or rightmost targets). As for human-human pairs, we also found that participants attained a group benefit together with the computer partner (Wahn & Kingstone, 2021).

To date, however, little research has explored if the coordination effort associated with labor division is related to physiological correlates of mental effort while physiological correlates in joint tasks of, for

---

* Corresponding author at: University of British Columbia, Psychology Department, 2136 West Mall, V6T 1Z4 Vancouver, Canada.
E-mail address: wahn@psychologie.uni-hannover.de (B. Wahn).

https://doi.org/10.1016/j.actpsy.2021.103291
Received 21 July 2020; Received in revised form 19 January 2021; Accepted 2 March 2021

instance, error processing (Czeszumski, Ehinger, Wahn, & König, 2019; Loehr, Kourtis, Vesper, Sebanz, & Knoblich, 2013), decision-making (Baumgart et al., 2020) or leader-follower dynamics (Konvalinka et al., 2014) have been explored. A physiological correlate that has often been linked to a variety of cognitive processes – one of them being mental effort – is pupil size (for a recent review, see Mathôt (2018)). Particularly for the MOT task, several studies have found that pupil size increases with an increasing number of tracked targets (Alnæs et al., 2014; Wahn, Ferris, Hairston, & König, 2016; Wardhani, Mathot, Boehler, & Laeng, 2019), suggesting that pupil size increases with mental effort.

In the present study, we aim to extend these earlier findings by examining if pupil size can detect an increase in mental effort assumed to be involved with coordination efforts. In particular, we ran two separate experiments (a main experiment and control experiment), in which each group of participants performed a MOT task jointly with a computer partner (as in our earlier study (Wahn & Kingstone, 2021)) and also alone. In the main experiment, participants were required to coordinate a labor division strategy with their computer partner. In the control experiment, no coordination effort was required as a pre-determined labor division was assigned to the participant. If pupil size is sensitive to a coordination effort, we expect that it will increase in the main experiment – relative to performing the MOT task alone – while this should not be the case for the control experiment.

2. Materials and methods

2.1. Participants

29 students (22 female, 7 male, $M = 19.28$ years, $SD = 1.28$ years) of the University of Groningen participated in the present study. 16 students participated in our main experiment (matching the number of participants in our earlier studies (Wahn et al., 2017; Wahn & Kingstone, 2021)) while the remaining 13 participants took part in our control experiment. Because of lab shutdown due to measures against the Corona virus, the intended sample size of 16 for the control experiment was not reached. Yet, for reasons that will be discussed later, we are confident that the sample size for the control experiment is still sufficient. The study was approved by the ethics review board of the University of Groningen (ECP approval code: PSY-1920-S-0254). All participants gave their consent in written form and received course credits for participation.

2.2. Experimental setup

Participants were seated in front of a computer screen (screen resolution: 1920 × 1080; 24″ monitor; 60 Hz refresh rate) at a 100 cm distance. Participants' head rested on a chinrest and a keyboard and computer mouse were placed in front of them within easy reach. Eye-tracking data was collected using an EyeTribe (30 Hz sampling rate). Eye position was calibrated using a 9-point calibration. The accuracy of the EyeTribe's measurements have been assessed for research purposes in a recent paper (Dalmaijer, 2014) and deemed sufficient for fixation analyses and pupilometry.

2.3. Experimental procedure

In the main experiment, participants performed a solo and joint condition in separate blocks of trials. In the solo condition (see top row in Fig. 1, for an example trial), participants first saw 19 stationary objects (0.36 visual degrees radius) on the computer screen. The locations of these objects were randomly selected. Either 0 ("non-tracking" trials) or 6 ("tracking" trials) objects turn white to indicate the target objects. The objects reverted to looking the same and started moving across the screen for 11 s. The objects' movement directions and speed (varying between 0.04 and 0.05 visual degrees per frame) were randomly chosen. Moreover, objects repelled each other in a physically plausible way (i.e, angle of incidence equaled the angle of reflection) and "bounced" of the screen borders. In the tracking trials, participants were required to track the movements of the target objects. In the non-tracking trials, participants were only required to look at the central dot (0.06 visual degrees radius). The purpose of the non-tracking trials was to replicate earlier results (Wahn, Ferris, et al., 2016) by assessing the relative increase in pupil sizes due to performing the MOT task. After objects stop moving, the trial was complete in the non-tracking trials. For the tracking trials, using the computer mouse, participants selected the objects that they thought were the targets. Participants could select as many objects as they wished, and confirmed their selection and ended the trial by clicking on the middle dot. Participants were instructed to be accurate in their selections and that each correct selection would count as one point

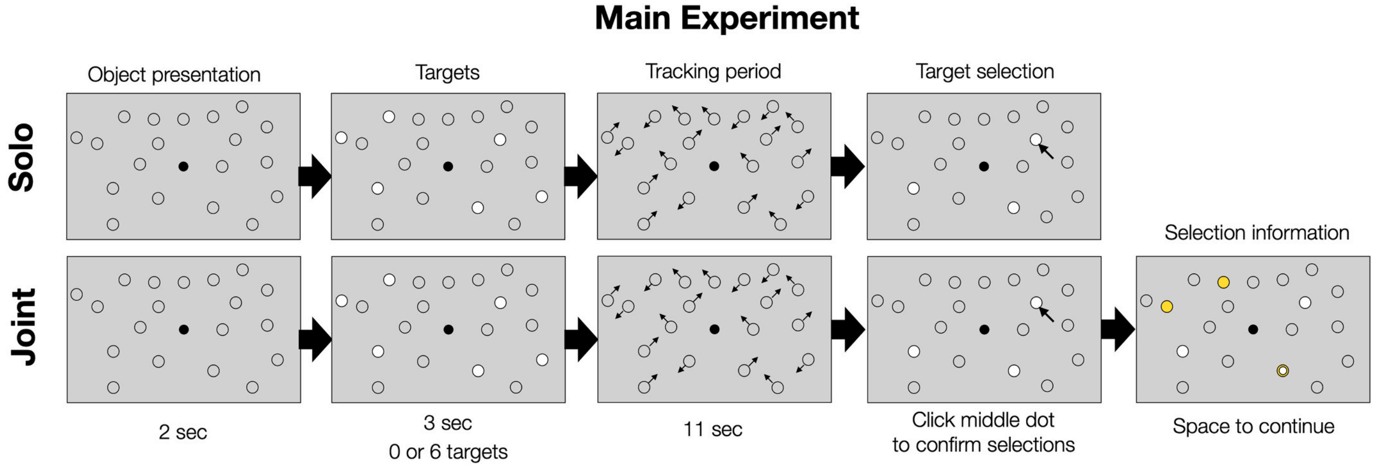

Fig. 1. Main experiment: Example trial sequence of a tracking trial, separately for the solo (top row) and joint (bottom row) conditions. Participants first saw 19 stationary objects. Then, a subset of six objects was indicated as target objects (in white) in the tracking trials. In the non-tracking trials, zero targets were indicated. During the tracking period, target indications were removed and objects moved across the screen. After the objects had stopped moving, participants could select the objects that they thought were targets, using the computer mouse, and subsequently confirmed their selections by clicking on the middle dot, which ended the trial for the solo condition. For the joint condition, after object selections, participants saw in yellow the object selections by the computer partner (in addition to their own selection shown in white). Overlapping selections were shown in yellow and white. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

toward their performance score, whereas each incorrect selection would result in one point being subtracted. No performance feedback was given. The solo condition contained a total of 25 trials (20 trials tracking trials; 5 non-tracking trials). Every fifth trial was a non-tracking trial. Participants knew that only in the tracking trials, targets were displayed. So it was evident at the beginning of a trial whether they would perform a tracking or non-tracking trial.

The sequence of events was the same for the joint condition as for the solo condition with the only change being that after participants confirmed their target selections, they would see in yellow the selections of the computer partner. Selections that overlapped between the participant and computer partner were shown in yellow and white (see last box, bottom part of Fig. 1). In earlier work on human-human pairs (Wahn et al., 2017) and on human-computer pairs (Wahn & Kingstone, 2021), this selection information was sufficient for participants to coordinate their actions to efficiently divide the labor.

For the joint condition, participants were instructed to work together with the computer partner by maximizing the number of correct target selections. They were also informed that overlapping correct selections would not count twice to their pair performance score and thus it is better to minimize the overlap in selections. As in our earlier study on human-computer pairs (Wahn & Kingstone, 2021) and prior to performing any trials, the computer partner was described in the following way: "The computer partner has been designed to behave in a human-like way. Specifically, we took data from human participants when they performed this tracking task together and used this data to model the behavior of the computer partner." The computer partner either consistently tracked the three leftmost or rightmost targets. The joint condition contained a total of 50 trials (40 tracking trials; 10 non-tracking trials). Again, every fifth trial was a non-tracking trial. Note that the trial number in the joint condition was doubled (relative to the solo condition) to make sure that participants have a sufficient number of trials to establish labor division strategies (Wahn et al., 2017; Wahn & Kingstone, 2021) and for the purpose of comparing the joint conditions across experiments for later trials (when labor division strategies are likely already formed). In particular, in our human-computer study (Wahn & Kingstone, 2021), pairs established devision of labor strategies relatively early in the experiment (i.e., within the first several trials).

In the control experiment, using a different group of participants, we only changed the joint condition by indicating to participants that they should track a certain set of targets when the targets were displayed at the beginning of a trial (i.e., either consistently the three leftmost or rightmost targets, referring to the positions of the targets at the start of the trial). Moreover, participants were informed that the computer partner would track the complementary set of targets. Thus, in the control experiment there is no coordination effort required to distribute the labor with the computer partner (for an example trial sequence, see Fig. 2).

For both experiments, the order of conditions was counterbalanced across participants. Note, however, that for the control experiment the number of participants for each order is not equal (8 for solo first; 5 for solo second) while for the main experiment we had an equal number of participants for each order (8 solo first; 8 solo second). Prior to running the experiment, participants performed a few trials for each condition to become familiar with the trial sequence. Each experiment took about 40 min to complete.

The experiments were programmed using Python 3.7.4. We ran them in OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) and used the PyGaze package (Dalmaijer, Mathôt, & Van der Stigchel, 2014) to integrate the EyeTribe within the experiment.

2.4. Methods of data analysis

2.4.1. Eyetracking data pre-processing

Prior to the analysis of pupil size, we removed all recorded samples from the eyetracking data, in which no data was recorded (e.g., due to blinks; 1% of samples removed for the main and control experiment, respectively). We then calculated the median pupil size between 3 and 8 s of the tracking period for each trial, matching our approach to the pupil size analysis in Wahn et al. (2016a). We chose this time window to avoid any influences on pupil sizes related to seeing the target indications at the beginning of a trial and motor preparation for the target selections toward the end of a trial. We also calculated the eye position deviation (in visual degrees) from the central dot for each sample and

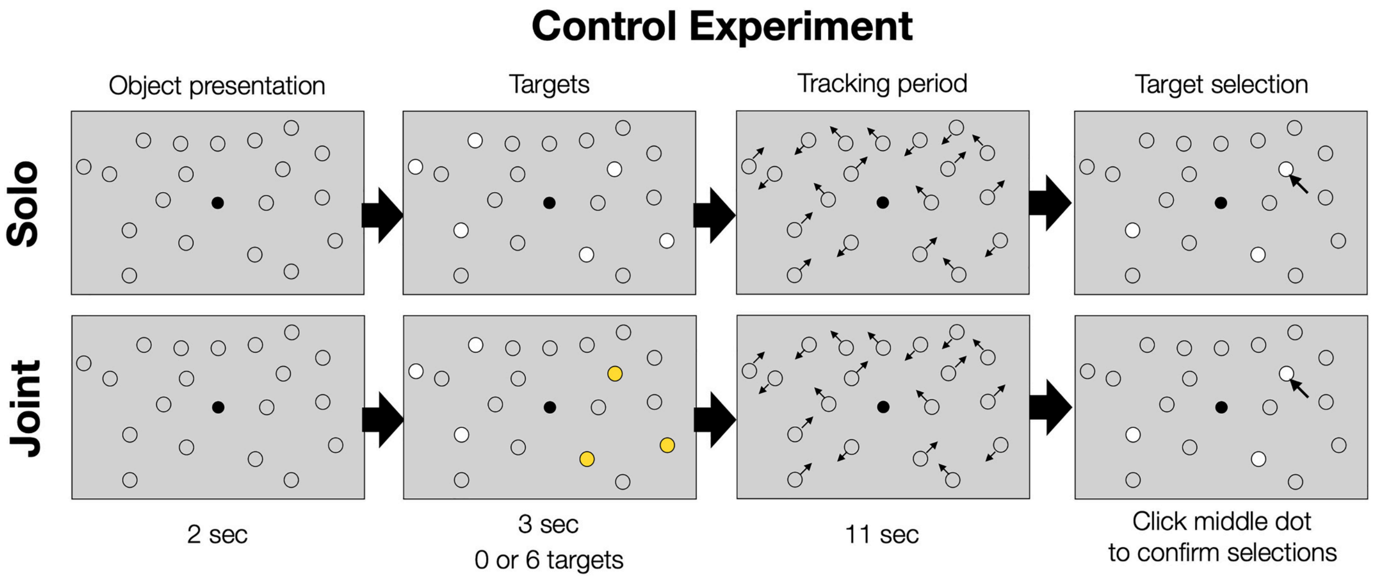

Fig. 2. Control experiment: Example trial sequence of a tracking trial, separately for the solo (top row) and joint (bottom row) conditions. Participants first saw 19 stationary objects. For the solo condition, a subset of six objects was indicated as target objects (in white) in the tracking trials. Note, in the non-tracking trials, zero targets were indicated. In the joint condition, the labor division was pre-determined: Three targets were assigned to the participant (shown in white), three targets to the computer partner (shown in yellow). In the tracking period, target indications were removed and objects moved across the screen. After objects stop moving, participants could select the objects that they believed to be the targets using the computer mouse. They confirmed their selections by clicking on the middle dot, which also ended the trial for both the solo and joint conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

took the median deviation across the extracted time window in the tracking period. We removed all trials, in which participants' deviation was more than 2 visual degrees from the center (4% of trials removed for the main experiment; zero trials removed for the control experiment). We used the median to have a more robust measure (with regard to outliers) to extract data for each trial. However, the median was only used to aggregate data across time within a single trial. For all subsequent analysis procedures (e.g., aggregating data across trials), the mean was used.

For our analyses, we matched the number of trials for the solo and joint condition (i.e., used trials of the first half of the joint condition) to avoid the potential confound that a longer exposure to the joint condition may have on pupil size. As noted above, the trials in the joint condition were doubled to allow comparisons between the joint conditions across experiments, for which we used all of the trials.

As the pupil sizes were recorded in an arbitrary unit (i.e., pixels) and hence it is difficult to adequately assess the magnitude of effects, we also provide the relative percentage increase in pupil sizes (between conditions) as a more intuitive measure of the reported effects as well as Cohen's *d* as an effect size measure for the effects of interest.

3. Results

3.1. Main experiment

As a replication of earlier results (Wahn, Ferris, et al., 2016), we assessed whether participants had larger pupil sizes in tracking trials compared to non-tracking trials, separately for the solo and joint conditions. For this purpose, we used the averaged pupil sizes across trials (for each participant and condition) and paired samples *t*-tests. For each comparison, we also computed percentage increase in pupil sizes in the tracking trials relative to the non-tracking trials. Pupil size significantly increased with tracking for the solo condition ($t(15) = 2.42$, $p = .023$, 2.91% increase, Cohen's $d = 0.60$, non-tracking trials $M = 18.05$ pixels ($SE = 0.36$), tracking trials $M = 18.56$ pixels ($SE = 0.36$)) and also for the joint condition ($t(15) = 3.54$, $p = .003$, 3.82% increase; Cohen's $d = 0.88$, non-tracking trials $M = 18.33$ pixels ($SE = 0.34$), tracking trials $M = 19.04$ pixels ($SE = 0.42$)). These findings replicate the relation between attentional load (operationalized as targets tracked in a MOT task) and pupil size (Alnæs et al., 2014; Wahn, Ferris, et al., 2016; Wardhani et al., 2019). For the rest of the main experiment analyses, we will consider the tracking trials.

Focusing now on the tracking trials, in order to address our research question of whether pupil size differs between solo and joint conditions, we ran a linear mixed model (Bates, Mächler, Bolker, & Walker, 2015) with the factors Condition (Solo, Joint) and Order (Solo first, Solo second), random intercepts for each participant, and pupil size (in pixels) as the dependent variable (for a descriptive overview of the Condition factor, see Fig. 3a). To assess significance, we used the Satterthwaite's method (Luke, 2017; Satterthwaite, 1941) and type III sum of squares. We did not add random slopes to the model as that would have led to a model convergence error. We find a significant main effect of Condition ($F(1,16) = 5.94$, $p = .027$) but no other significant effects (main effect of Order: $F(1,16) = 2.26$, $p = .152$; interaction effect: $F(1,16) = 0.85$, $p = .370$). For the main effect of Condition, the pupil size increase in the joint condition ($M = 19.04$ pixels, $SE = 0.42$) relative to the solo condition ($M = 18.56$ pixels, $SE = 0.36$) is (on average) 2.63% (Cohen's $d = 0.57$, *M difference* $= 0.48$ pixels, $SE = 0.21$), which is similar to the increase in pupil size we found above when comparing the non-tracking trials and tracking trials in the solo condition (2.91%), and in general similar to the size of cognitively driven pupil size changes (Mathôt, 2018). In short, pupil size was larger for the joint condition than for the solo condition.

As a control analysis, we ran the same model with the eye deviation (from the center) added as a covariate and obtained the same results. We again found a significant main effect of Condition ($F(1,16.85) = 5.18$, $p = .036$) and no other significant effects (main effect of Order: $F(1,15.90) = 1.99$, $p = .178$; interaction effect: $F(1,20.21) = 0.34$, $p = .565$; effect of eye deviation: $F(1,30.26) = 0.32$, $p = .577$).

Given that pupil size increased in the joint condition relative to the solo condition, we next tested whether participants actually distributed the labor in the joint condition. To quantify the efficiency of how well the labor is distributed, we assessed for each trial how many objects were selected both by the computer and the participant ("overlapping selection"). Ideally, this number should be low if a participant and the computer partner divided the labor well. We then divided the number of overlapping selections by the total number of selections for each trial to obtain the fraction of overlapping selections (henceforth referred to as "overlap"). For this measure, a low score would indicate an efficient labor division whereas a high score would indicate a non-efficient labor division. To have a baseline to which compare the overlap, we calculated the expected overlap for the solo condition by simulating the selections the computer partner would take in this condition. The rationale for this baseline is that the computer partner's behavior is the same for the solo and joint condition and that a reduction in overlap would only be present due to a change in the participants' behavior.

We ran a linear mixed model with the factors Condition (Baseline, Joint) and Order (Solo first, Solo second), random intercepts for all

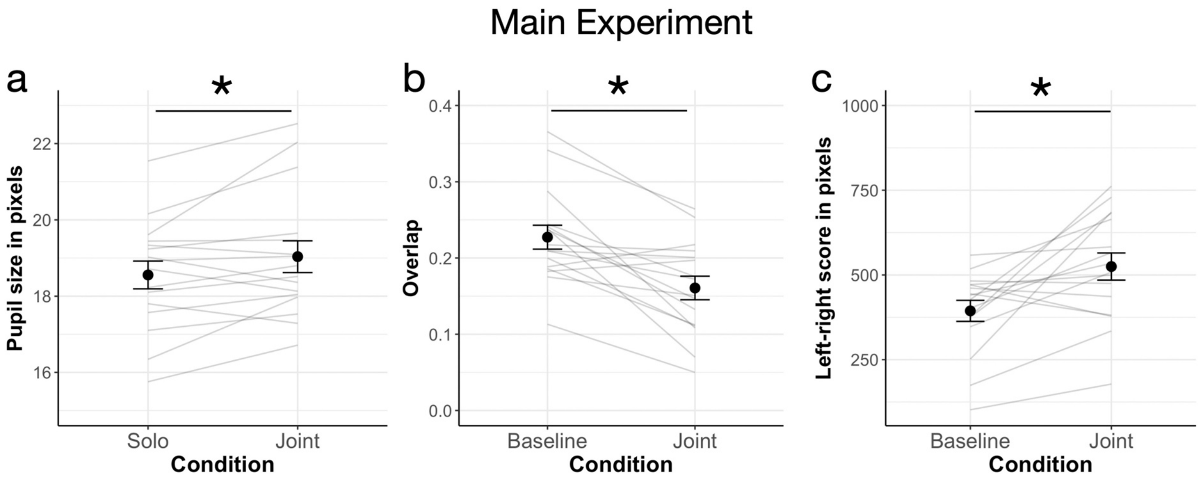

Fig. 3. Main experiment results: Averages across participants are displayed for (a) pupil size, (b) overlap (i.e., the fraction of overlapping selections relative to the total number of selections), and (c) left-right scores, separately of the levels of the within-subject factor Condition. The gray lines indicate the individual participants' data. Error bars in all panels are standard error of the mean.

participants, and the dependent variable overlap. We found a significant main effect of Condition ($F(1,16) = 5.21$, $p = .037$) but no other significant effects (main effect of Order: $F(1,24.46) = 0.16$, $p = .684$; interaction effect: $F(1,16) = 0.70$, $p = .416$). These results indicate that participants distributed the labor in the joint condition (for a descriptive overview, see Fig. 3b), replicating our earlier results (Wahn & Kingstone, 2021).

While the analysis above confirmed that participants distributed the labor, we also verified whether this overlap reduction was attained by a left and right labor division. That is, one possibility is that participants' reduction in the overlap could have been attained by simply trying to avoid the selections of the computer partner (without a clear sense whether they follow a certain pattern or not). For this purpose, we computed a measure quantifying the extent to which participants used such a labor division. For this measure, we first extracted for each trial the initial horizontal starting positions (in pixels) of all correct object selections by the participant and computer partner. We then averaged across these positions for each pair member and computed the absolute difference between the participant and computer partner averages. The resulting values (henceforth referred to as "left-right scores") indicate to what extent pair members used a left and right labor division. As for the overlap, we compared the left-right scores to the hypothetical left-right scores pair members would have attained in the solo condition (for a descriptive overview, see Fig. 3c).

We again ran a linear mixed model with the factors Condition, Order, random intercepts for all participants, and the left-right scores as the dependent variable. We found a significant main effect of Condition ($F(1,16) = 9.22$, $p = .009$) but no other significant effects (main effect of Order: $F(1,23.67) = 0.82$, $p = .374$; interaction effect: $F(1,16) = 3.76$, $p = .070$). Finally, we correlated the left-right scores with the overlap, returning a significant negative correlation ($r = -.90$, $p < .001$). These findings indicate that participants used a left and right labor division.

Given that participants clearly divided the labor with the computer partner, we asked whether this labor division was actually beneficial in a sense that the human-computer pairs in the joint condition outperformed participants in the solo condition. To answer this question, we ran a linear mixed model with the factors Condition, Order, random intercepts for all participants, and the performance as the dependent variable. Importantly, for the joint condition we used the pair performance and for the solo condition the individual performance as the dependent variable. We found a significant main effect of Condition ($F(1,16) = 128.65$, $p < .001$), suggesting that human-computer pairs clearly outperformed individual humans in the solo condition. Moreover, we found a significant interaction effect ($F(1,16) = 24.07$, $p < .001$) while the Order effect was not significant ($F(1,16) = 0.476$, $p = .500$). The significant interaction effect was driven by a larger performance advantage for the joint condition relative to the solo condition when the solo condition was performed first (solo: $M = 1.71$ points ($SE = 0.25$ vs. joint: $M = 3.88$ points ($SE = 0.23$) compared to when the solo condition was performed second (solo: $M = 2.31$ points ($SE = 0.14$) vs. joint: $M = 3.66$ points ($SE = 0.21$).

3.2. Control experiment

As above for the main experiment and as a replication of earlier results (Wahn, Ferris, et al., 2016), we first assessed if participants had larger pupil sizes in tracking trials compared to non-tracking trials, separately for each condition, using the averaged pupil sizes across trials (for each participant and condition) and paired samples t-tests. Again, for each comparison we also computed the percentage increase in pupil sizes in the tracking trials relative to the non-tracking trials. Also for the control experiment, pupil size significantly increased with tracking for the solo condition ($t(12) = 5.04$, $p < .001$, 4.12% increase, Cohen's $d = 1.40$, non-tracking trials: $M = 16.28$ pixels ($SE = 0.44$), tracking trials: $M = 16.97$ pixels ($SE = 0.55$)) and for for the joint condition ($t(12) = 3.33$, $p = .006$, 2.30% increase, Cohen's $d = 0.92$, non-tracking trials: $M = 16.61$ pixels ($SE = 0.42$), tracking trials: $M = 17.01$ pixels ($SE = 0.49$)). When comparing these percentages in pupil size increase across experiments, separately for the solo and joint conditions, we found no significant differences (solo condition: $t(27) = -0.81$, $p = .423$; joint condition: $t(27) = 1.14$, $p = .262$).

With the control experiment, we assessed whether the increase in pupil size in the joint condition (relative to the solo condition) found in the main experiment is specifically related to a coordination effort associated with implementing division of labor strategies. As a reminder, in the control experiment division of labor strategies were pre-determined. If the increase in pupil sizes in the joint condition (relative to the solo condition) in the main experiment is specifically related to participants' coordination effort to devise labor division strategies, then a pupil size increase should be absent in the control experiment.

We ran the same linear mixed model as before with pupil size as the dependent variable and the factors Condition and Order and random intercepts for all participants. We no longer find a significant main effect of Condition ($F(1,13) = 0.55$, $p = .471$) nor any other significant effects (Order: $F(1,13) = 0.10$, $p = .755$; interaction effect: $F(1,13) = 4.29$, $p = .059$). We also computed a Bayes factor for the Condition effect (Cohen's $d = 0.06$) and found that the data is 3.51 more likely under the null hypothesis than under the alternative hypothesis (for a descriptive overview, see Fig. 4a), which can be considered "substantial evidence" for the null hypothesis (Wetzels et al., 2011). For computing the Bayes factor, we used the BayesFactor package in R, which uses a Jeffreys prior for the variance and a Cauchy prior for the standardized effect size (Rouder, Morey, Speckman, & Province, 2012). This clear null finding for the control experiment suggests that the difference in statistical power for the control and main experiments due to the different sample sizes is not an issue. In sum, these findings suggest that for the control experiment pupil size is not increased in the joint condition (relative to the solo condition).

To verify that participants followed the instructions to track the indicated targets, we computed the overlap and the left-right scores for the control experiment as well. Running the same linear mixed models as for the main experiment, we find for the overlap a significant main effect of Condition ($F(1,26) = 21.05$, $p < .001$) and no other significant effects (Order: $F(1,26) = 0.01$, $p = .930$; interaction effect: $F(1,26) = 0.00$, $p = .972$) – for a descriptive overview, see Fig. 4b. The same is also true for the left-right scores: We find a significant main effect of Condition ($F(1,26) = 5.56$, $p = .026$) and no other significant effects (Order: $F(1,26) = 0.40$, $p = .531$; interaction effect: $F(1,26) = 0.118$, $p = 0.734$) – for a descriptive overview, see Fig. 4c. In short, participants followed the task instructions in the control experiment by tracking the indicated targets.

3.3. Comparisons across experiments

While the results suggest that differences in pupil size in the main experiment are related to a coordination effort to implement labor division strategies, an open question is whether the percentage increase in pupil size between the solo and joint condition is also larger in the main experiment compared to the control experiment. Comparing these increases between experiments using an independent t-test, we only find on a descriptive level (main: 2.63% vs. control: 0.42%) support for this comparison but it did not reach significance ($t(27) = 1.57$, $p = .128$, Cohen's $d = 0.59$).

In addition, one could suggest that the increase in pupil size in the main experiment could simply be driven by an increase in the individual effort by the participants to track targets. In other words, participants may have tried to track more targets in the joint condition in the main experiment relative to the solo condition, which could alternatively explain an increase in pupil size in the joint condition compared to the solo condition. Conversely, given that the number of tracked targets was already pre-determined, such a (potential) increase in the individual

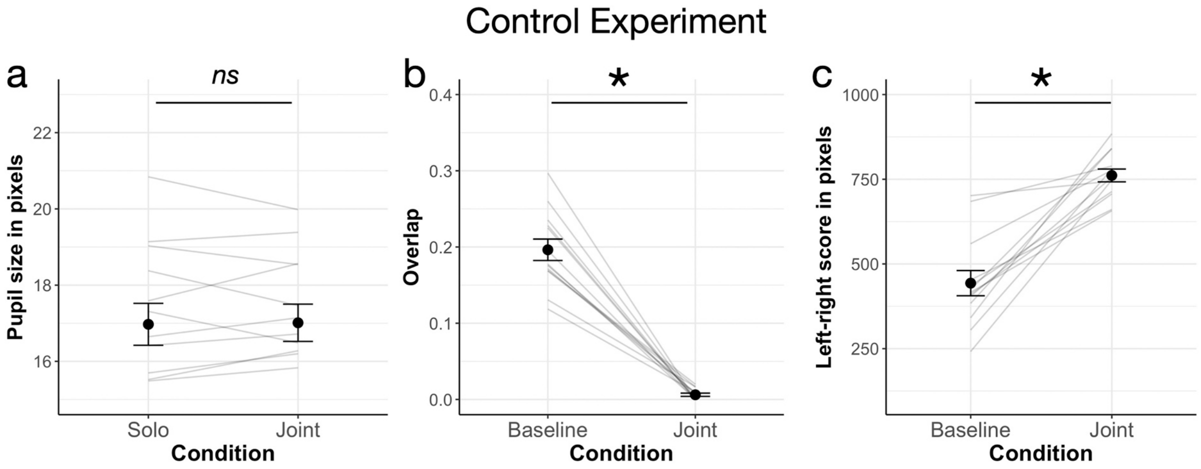

Fig. 4. Control experiment results: Averages across participants are displayed for (a) pupil size, (b) overlap (i.e., the fraction of overlapping selections relative to the total number of selections), and (c) left-right scores, separately of the levels of the within-subject factor Condition. The gray lines indicate the individual participants' data. Error bars in all panels are standard error of the mean.

effort may not have been present in the control experiment. A different point is that we also have not addressed whether there are time-on-task effects (i.e., training or fatigue effects) with regard to the individual performance of participants. To address this alternative explanation of our pupil size results and potential time-on-task effects, we ran a linear mixed model with the factors Condition, Order, and Experiment, random intercepts, and participants' *individual* performance as the dependent variable. For this model, neither the main effects (Condition: $F(1,29) = 0.12, p = .730$; Order: $F(1,29) = 2.05, p = .163$; Experiment: $F(1,29) = 1.54, p = .225$) nor interaction effects were significant (Condition x Order: $F(1,29) = 2.40, p = .132$; Condition x Experiment: $F(1,29) = 2.41, p = .131$; Order x Experiment: $F(1,29) = 1.38, p = .250$; Condition x Order x Experiment: $F(1,29) = 1.67, p = .288$). These findings suggest that the participants' individual effort to track her/his own targets stayed relatively constant across conditions. Taken together with our results from the main experiment, our findings suggest that the differences in pupil size for the main experiment between the joint and individual condition are likely due to a rise in mental effort due to coordinating labor with the computer partner. In addition, given there are no significant Order effects (or interaction effects involving the factor Order), our findings suggest that there are no time-on-task effects.

Another open question is whether this rise in effort to divide the labor leads to a comparable overlap and pair performances relative to the control experiment, where participants already were instructed to use a perfect labor division. To address this question, we directly compared the overlap and pair performances between experiments. Given that participants may still devise labor division strategies in the first half of trials in the joint condition, we ran linear mixed models that not only include the between-subject factor Experiment (Main, Control) but also the within-subject factor Half (1st, 2nd) and we again included random intercepts for participants. The dependent variables are the pair and overlap performances, respectively (for a descriptive overview, see Fig. 5).

For the pair performances, we find a significant main effect of Experiment ($F(1,58) = 4.02, p = .049$) but no other significant effects (Half: $F(1,29) = 1.31, p = .262$; Experiment x Half: $F(1,29) = 0.31, p < .580$). The same is also true for the overlap. We find a significant main effect of Experiment ($F(1,57.88) = 31.36, p < .001$) but no other significant effects (Half: $F(1,29) = 0.11, p = .742$; interaction effect: $F(1,29) = 0.02, p = .876$). In sum, the participants' division of labor strategies in the main experiment result in a lower pair performance (and higher overlap) than the pre-defined strategies in the control experiment, suggesting that participants' labor division strategies do not reach the levels of efficiency as in the control experiment.

4. Discussion

In the present study, we investigated whether pupil size – a well-established physiological correlate of mental effort – would increase with coordination effort. In particular, we tested whether the effort

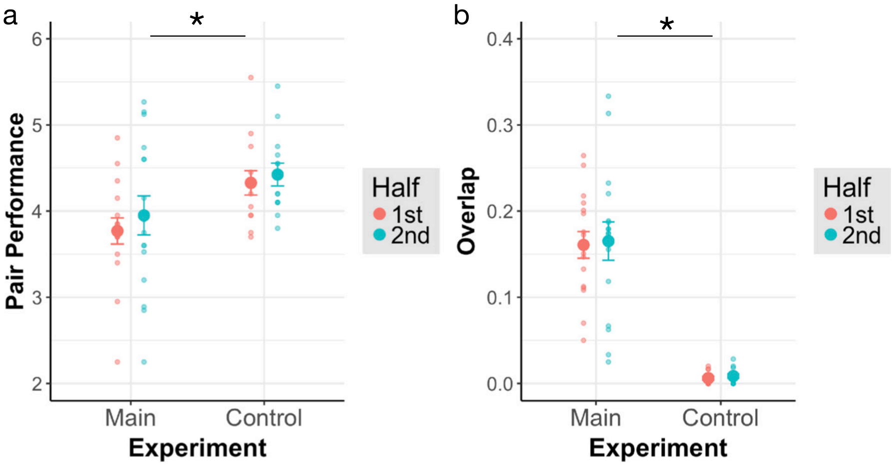

Fig. 5. Comparisons of joint conditions across experiments: Averaged performances across participants for (a) pair performance and (b) overlap, separately for the main and control experiment. The pair performances are the combined scored points from the participant and computer partner for each trial. In particular, each correct selection adds one point to the performance; each incorrect selection subtracts one point. Importantly, overlapping correct selections were counted only once to the pair performance. The overlap is computed by dividing the overlapping selections by the total number of selections for each trial. Error bars in all panels are standard error of the mean.

associated with implementing labor division strategies in a joint MOT task would result in a pupil size increase relative to a condition, in which the MOT task is performed alone. In our main experiment, we found that pupil size increased by about 3% when a coordination effort was required relative to when the MOT task was performed alone. The magnitude of this effect is comparable to other cognitively driven changes in pupil sizes such as the effect of tracking several objects in a MOT task (Wahn, Ferris, et al., 2016) or holding several items in working memory (Mathôt, 2018). Importantly, we did not find this pupil size increase in a control experiment, in which no coordination effort was required. These results indicate that pupil size increases with mental effort associated with coordinating labor division in a joint visuospatial task. On top of that, we also replicate earlier findings showing that pupil size increases with attentional load in a MOT task (Alnæs et al., 2014; Wahn, Ferris, et al., 2016; Wardhani et al., 2019), that participants are willing to divide labor with a computer partner, and that this labor division is clearly beneficial, resulting in a group benefit (Wahn & Kingstone, 2021).

Previous work (for a recent review, see Mathôt (2020)) has linked changes in pupil size to a variety of different types of mental effort (e.g., memory or attentional load), which suggest that changes in pupil size may reflect a general increase in mental effort, regardless of the cognitively demanding task. The present findings lend strong support to this view by showing that changes in pupil size are also sensitive to a coordination effort.

We propose that the coordination effort indexed by a pupil size increase likely represents the combined effort related to several coordination mechanisms that are required to successfully coordinate with others. To split this combined effort into its constituent parts, we suggest that one crucial aspect of it is uncertainty about the actions of the computer partner, or, in other words, the effort related to deciphering the “intentions” of the computer partner (i.e., that it consistently chooses the leftmost or rightmost targets). Relatedly, previous work on pupil sizes and uncertainty indeed has shown that pupil sizes increase with uncertainty (Richer & Beatty, 1987; Urai, Braun, & Donner, 2017) and hence the uncertainty about the computer’s actions could (at least in part) drive the present effects. Another component is the effort related to adapting one’s own behavior to the computer’s actions to facilitate coordination (i.e., by tracking a complementary set of targets). A future study could aim to disentangle these two components to assess their individual and combined effects on pupil size.

In the present study, the coordination effort was one-sided because the computer’s behavior remained consistent throughout. As a result, all the required coordination effort was carried out by the participant. We know from human-human pairs (Wahn et al., 2017) that such one-sided coordination efforts are not unusual. That is, there are several instances where one co-actor in a pair does not deviate from her/his own tracking behavior and only the other co-actor adapts her/his actions by tracking a complementary set of targets. Thus, the present study investigated (to a degree) coordination behavior that could also occur in human-human pairs. Nonetheless, it would be interesting to investigate in future studies whether pupil size increases to a lesser degree if both co-actors are exerting efforts to coordinate task load, thus distributing the required effort across *both* co-actors.

With regard to other joint tasks, coordination efforts may also involve other cognitive processes such as, for instance, co-representing (Schmitz, Vesper, Sebanz, & Knoblich, 2018) or predicting (Konvalinka, Vuust, Roepstorff, & Frith, 2010) the co-actor’s actions – for a recent review on cognitive mechanisms in joint tasks, see Vesper et al. (2017). Relatedly, labor coordination may also occur spontaneously as an emergent phenomenon (Richardson et al., 2015) rather than as a planned coordination (as in the present study). Therefore it is an open question whether such emergent labor divisions also require a coordination effort that is reflected in pupil size. Another consideration is that the coordination of labor could also occur more in real-time (Benerink, Zaal, Casanova, Bonnardel, & Bootsma, 2016, 2018) than in the present study. As noted above, part of the participants’ coordination effort in the current study involved deciphering the “intentions” of the computer partner and then adapting their behavior accordingly for the subsequent trial(s). In contrast, such action adaptions may also occur in real-time (i.e. participants directly react to, or anticipate, the actions of their partner). For such more real-time joint tasks, coordination efforts likely reflect other coordination mechanisms, and it is again an open question if pupil size is also sensitive to such real-time coordination efforts. Taking a different tack, labor divisions can also occur in other cognitive domains such as memory (Bietti & Sutton, 2015; Rajaram & Pereira-Pasarin, 2010), problem-solving (Laughlin, Bonner, & Miner, 2002; Laughlin, Hatch, Silver, & Boh, 2006; Roberts & Goldstone, 2011), decision-making (Kerr & Tindale, 2004), and visuomotor processing (Knoblich & Jordan, 2003; Newman-Norlund, Bosga, Meulenbroek, & Bekkering, 2008; Reed et al., 2006; Ganesh et al., 2014; Van der Wel, Knoblich, & Sebanz, 2011; Takagi, Ganesh, Yoshioka, Kawato, & Burdet, 2017; Wahn, Schmitz, König, & Knoblich, 2016; Wahn, Karlinsky, Schmitz, & König, 2018). For several of these domains, the involved cognitive effort has been related to changes in pupil size (for a review, see: Mathôt (2020)). However, it has not been tested whether pupil size will also be sensitive to a *coordination* effort in such tasks as well. Taken together, addressing these questions in future studies will reveal if changes to pupil size are also sensitive to a wide variety of coordination efforts across joint tasks as it follows that, depending on the task, other cognitive processes will surely be required for coordinating actions with others. Finally, from a more technical perspective, a future study could also test whether it would be possible to use pupil size to determine (Mathôt, Melmi, Van Der Linden, & Van der Stigchel, 2016) if coordination strategies are ongoing (e.g., being formulated) or not (e.g., have been established).

CRediT authorship contribution statement

Basil Wahn, Alan Kingstone, & Sebastiaan Mathôt: Funding acquisition, Conceptualization, Writing - reviewing & editing. Basil Wahn, Veera Ruuskanen, & Sebastiaan Mathôt: Methodology. Veera

Ruuskanen: Data collection. Basil Wahn: Software, Visualization, Analysis, Writing - original draft.

Sebastiaan Mathôt & Alan Kingstone: Supervision.

Acknowledgments

We acknowledge the support of a DFG research fellowship (WA 4153/2-1) awarded to BW, and an NSERC Discovery Grant awarded to AK.

References

Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision, 14*(4), 1–20.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Baumgart, K. G., Byvshev, P., Sliby, A.-N., Strube, A., König, P., & Wahn, B. (2020). Neurophysiological correlates of collective perceptual decision-making. *European Journal of Neuroscience, 51*(7), 1676–1696.

Benerink, N. H., Zaal, F. T., Casanova, R., Bonnardel, N., & Bootsma, R. J. (2016). Playing “pong” together: Emergent coordination in a doubles interception task. *Frontiers in Psychology, 7*, 1910.

Benerink, N. H., Zaal, F. T., Casanova, R., Bonnardel, N., & Bootsma, R. J. (2018). Division of labor as an emergent phenomenon of social coordination: The example of playing doubles-pong. *Human Movement Science, 57*, 134–148.

Bietti, L. M., & Sutton, J. (2015). Interacting to remember at multiple timescales: Coordination, collaboration, cooperation and culture in joint remembering. *Interaction Studies, 16*(3), 419–450.

Brennan, A. A., & Enns, J. T. (2015a). What’s in a friendship? Partner visibility supports cognitive collaboration between friends. *PLoS One, 10*(11), Article e0143469.

Brennan, A. A., & Enns, J. T. (2015b). When two heads are better than one: Interactive versus independent benefits of collaborative cognition. *Psychonomic Bulletin & Review, 22*(4), 1076–1082.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*(3), 1465–1477.

Czeszumski, A., Ehinger, B. V., Wahn, B., & König, P. (2019). The social situation affects how we process feedback about our actions. *Frontiers in Psychology, 10*, 361.

Dalmaijer, E. (2014). *Is the low-cost eyetribe eye tracker any good for research?* PeerJ PrePrints: Technical report.

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). Pygaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods, 46*(4), 913–921.

Ganesh, G., Takagi, A., Osu, R., Yoshioka, T., Kawato, M., & Burdet, E. (2014). Two is better than one: Physical interactions improve motor performance in humans. *Scientific Reports, 4*.

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology, 55*, 623–655.

Knoblich, G., & Jordan, J. S. (2003). Action coordination in groups and individuals: Learning anticipatory control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(5), 1006–1016.

Konvalinka, I., Bauer, M., Stahlhut, C., Hansen, L. K., Roepstorff, A., & Frith, C. D. (2014). Frontal alpha oscillations distinguish leaders from followers: Multivariate decoding of mutually interacting brains. *Neuroimage, 94*, 79–88.

Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: Continuous mutual prediction and adaptation in joint tapping. *Quarterly Journal of Experimental Psychology, 63*(11), 2220–2230.

Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes, 88*(2), 605–620.

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: Effects of group size. *Journal of Personality and Social Psychology, 90*(4), 644–651.

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience, 25*(7), 1049–1061.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in r. *Behavior Research Methods, 49*(4), 1494–1502.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition, 1*(1).

Mathôt, S. (2020). Tuning the senses: How the pupil shapes vision at the earliest stage. *Annual Review of Vision Science, 6*.

Mathôt, S., Melmi, J.-B., Van Der Linden, L., & Van der Stigchel, S. (2016). The mind-writing pupil: A human-computer interface based on decoding of covert attention through pupillometry. *PLoS One, 11*(2).

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324.

Newman-Norlund, R. D., Bosga, J., Meulenbroek, R. G., & Bekkering, H. (2008). Anatomical substrates of cooperative joint-action in a continuous motor task: Virtual lifting and balancing. *Neuroimage, 41*(1), 169–177.

Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science, 5*(6), 649–663.

Reed, K., Peshkin, M., Hartmann, M. J., Grabowecky, M., Patton, J., & Vishton, P. M. (2006). Haptically linked dyads are two motor-control systems better than one? *Psychological Science, 17*(5), 365–366.

Richardson, M. J., Harrison, S. J., Kallen, R. W., Walton, A., Eiler, B. A., Saltzman, E., & Schmidt, R. (2015). Self-organized complementary joint action: Behavioral dynamics of an interpersonal collision-avoidance task. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 665.

Richer, F., & Beatty, J. (1987). Contrasting effects of response uncertainty on the task-evoked pupillary response and reaction time. *Psychophysiology, 24*(3), 258–262.

Roberts, M. E., & Goldstone, R. L. (2011). Adaptive group coordination and role differentiation. *PLoS One, 6*(7), Article e22377.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default bayes factors for anova designs. *Journal of Mathematical Psychology, 56*(5), 356–374.

Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika, 6*(5), 309–316.

Schmitz, L., Vesper, C., Sebanz, N., & Knoblich, G. (2018). Co-actors represent the order of each other's actions. *Cognition, 181*, 65–79.

Szymanski, C., Pesquita, A., Brennan, A. A., Perdikis, D., Enns, J. T., Brick, T. R., … Lindenberger, U. (2017). Teams on the same wavelength perform better: Inter-brain phase synchronization constitutes a neural substrate for social facilitation. *Neuroimage, 152*, 425–436.

Takagi, A., Ganesh, G., Yoshioka, T., Kawato, M., & Burdet, E. (2017). Physically interacting individuals estimate the partner's goal to enhance their movements. *Nature Human Behaviour, 1*, 1–6.

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications, 8*(1), 1–11.

Van der Wel, R. P., Knoblich, G., & Sebanz, N. (2011). Let the force be with us: Dyads exploit haptic coupling for coordination. *Journal of Experimental Psychology: Human Perception and Performance, 37*(5), 1420–1431.

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., Hristova, D., Karlinsky, A., McEllin, L., Nijssen, S., Schmitz, L., & Wahn, B. (2017). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology, 7*, 2039.

Wahn, B., Czeszumski, A., Labusch, M., Kingstone, A., & König, P. (2020). Dyadic and triadic search: Benefits, costs, and predictors of group performance. *Attention, Perception, & Psychophysics, pages*, 1–19.

Wahn, B., Ferris, D. P., Hairston, W. D., & König, P. (2016). Pupil sizes scale with attentional load and task experience in a multiple object tracking task. *PLoS One, 11* (12).

Wahn, B., Karlinsky, A., Schmitz, L., & König, P. (2018). Let's move it together: A review of group benefits in joint object control. *Frontiers in Psychology, 9*.

Wahn, B., & Kingstone, A. (2020). Labor division in joint tasks: Humans maximize use of their individual attentional capacities. *Attention, Perception, & Psychophysics, pages*, 1–11.

Wahn, B., & Kingstone, A. (2021). Humans share task load with a computer partner if (they believe that) it acts human-like. *Acta Psychologica, 212*, 103205.

Wahn, B., Kingstone, A., & König, P. (2017). Two trackers are better than one: Information about the co-actor's actions and performance scores contribute to the collective benefit in a joint visuospatial task. *Frontiers in Psychology, 8*, 669.

Wahn, B., Kingstone, A., & König, P. (2018). Group benefits in joint perceptual tasks – A review. *Annals of the New York Academy of Sciences, 1426*(1), 166–178.

Wahn, B., König, P., & Kingstone, A. (2020). Collaborative multiple object tracking: Benefits and predictors of group performance. *PsyArXiv [Preprint]*.

Wahn, B., Schmitz, L., König, P., and Knoblich, G. (2016). Benefiting from being alike: Interindividual skill differences predict collective benefit in joint object control. In *Proceedings of the 38th annual conference of the cognitive science society*, pages 2747–2752.

Wardhani, I. K., Mathot, S., Boehler, C. N., & Laeng, B. (2019). Effects of nicotine on pupil size and cognitive performance amongst non-nicotine users. *PsyArxiv [Preprint]., 158*, 45–55.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*(3), 291–298.