Sleep restriction increases coordination failure

Marco Castillo[a] and David L. Dickinson[b]

[a]Texas A&M University; Melbourne Institute, University of Melbourne; IZA (Bonn, Germany)
Castillo: marco.castillo@tamu.edu

[b]Appalachian State University; ESI (Chapman University); IZA (Bonn, Germany)
Dickinson: dickinsondl@appstate.edu (corresponding author)

April 21 2022

Abstract

When group outcomes depend on minimal effort (e.g., disease containment, work teams, or group hunt success), a classic coordination problem exists. Using a well-established paradigm, we examine how a common cognitive state (insufficient sleep) impacts coordination outcomes. Our data indicate that insufficient sleep increases coordination failure costs, which suggests that the sleep or, more generally, cognitive composition of a group might determine its ability to escape from a trap of costly miscoordination and wasted cooperative efforts. These findings are first evidence of the potentially large externality of a commonly experienced biological state (insufficient sleep) that has infiltrated many societies.

1 Introduction

Coordination games have widespread applications of interest across disciplines like economics, organizational behavior, and psychology. As such, how individuals solve coordination problems (or factors that predict coordination failure) are a natural focal point for behavioural research. Examples of coordination problems are found in a variety of diverse environments such as the occupational settings of team production or industrial disaster risk management,[1] the behavioral anthropology of aboriginal subsistence whaling, and containment efforts of communicable disease outbreaks, to name a few. In fact, the recent COVID 19 pandemic has made clear that stakes can be high in coordination environments where outcomes are dictated by the minimal effort within a group (e.g., family unit, social circle, larger community groups).

Here we extend our understanding of how deliberation (i.e., high-level cognition) affects coordination success by randomly inducing a common cognitive state prior to decision making—we manipulate participants' sleep levels to approximate either recommended nightly sleep levels or insufficient sleep levels experienced by a significant portion of the adult population in many countries (Hafner et al., 2017).[2] Importantly, in coordination settings where public health is at stake in very high profile ways in recent times (e.g., riot prevention efforts, COVID 19 first-respondent and medical care activities), critical workers who must coordinate activities are often the most likely to be sleep deprived.[3]

Game theorists also have a significant interest in coordination problems, in general, as a standard paradigm to study cooperative dilemmas. Because coordination games present a multiplicity of Pareto-ranked Nash equilibria, they contrast with other well-known cooperative dilemmas like the Prisoners Dilemma or common pool resource problem. And, some research suggests that individuals may coordinate on the most inefficient possible Nash outcome in such games (Cooper et al., 1992; Ochs, 1995; Van Huyck et al., 1990, 1991), which makes the identification of factors that lead to both coordination failure and inferior coordination outcomes of great importance. Our behavioural focus on sleep restriction as one such potential factor is intended as a highly real-world relevant way to manipulate the likelihood that participants use more automatic versus deliberative decision processes in making coordination choices.[4]

---

[1] Other examples of interest to organizational researchers have been noted in the literature (Knez and Camerer, 2000).

[2] Another approach in the literature to manipulate deliberation has been to impose a time constraint on making decisions. A recent paper studying a two-person coordination game (the Stag hunt game) reported improved coordination under time pressure, and the authors use this to suggest that deliberation may detract from coordination at least in certain contexts (Belloc et al., 2019).

[3] See Pirrallo et al. (2012). Also, the Walter Reed Army Institute of Research has noted the prevalence of insufficient sleep among emergency service and healthcare providers in recent efforts to support more successful COVID-19 first respondent activities.

[4] Killgore et al. (2012) highlights how decisions relying on critical components of the prefrontal cortex are

This paper contributes to the literature in a timely fashion. Using an ecologically valid protocol, we manipulated sleep to levels that are commonplace in modern society–nearly 30% of U.S. adults operate daily at the levels of sleep restriction (SR) we induced in our study (Schoenborn and Adams, 2010). Recent estimates using data from several industrialized countries found that these levels of insufficient sleep can cost an economy anywhere from 1%-3% of its annual GDP (Hafner et al., 2017).[5] Yet, little is known about how commonly experienced SR impacts group interactions or affects coordination efforts. Because SR can be thought of as an externally valid way to alter the cognitive mechanism used, our results will have implications for our understanding of the general underpinnings of decision making as well. And, the controlled decision environment we examined allows us to quantify the costs of coordination failure in a way that is difficult in naturally occurring field settings. Regarding hypotheses, we note in the following section that different mechanisms may suggest either increased or decreased effort choices in coordination settings when sleepy, and so our paper can be considered more exploratory. We find that SR tends to slow convergence to the equilibrium selection prediction in our environment, which can increase earnings variation in a way that depends on how many SR members are in the group. In an environment of repeated interaction with the same group members, which has been shown to improve coordination on Pareto superior Nash equilibria, our results show that SR members within the group can entirely eliminate the increased likelihood of successful coordination that comes with repeated interaction. Importantly, our data also reveal nontrivial costs associated with wasteful miscoordination, and these wasted effort costs increase when SR individuals are part of the group.

2 Background

Experimental research has well-established results showing coordination failure (Cooper et al., 1990, 1992; Van Huyck et al., 1990, 1991), and increased deliberative-thought capacity has been implicated in improved coordination (Mizrahi et al., 2020). While the behavioral literature is at times mixed on what factors most regularly improve coordination, it appears to be improved via shared experience of group members through repeated interactions or with communication (see Devetag and Ortmann (2007) for a review of these and other factors that have been found to impact outcomes in coordination games). While we do not examine communication in our study, we do exogenously vary the matching protocol (i.e., shared experience) across two treatments to examine outcome differences with random versus fixed group matching. Our evidence shows that repeated group interaction increases the likelihood of successful coordination in a 3-person minimum effort settings, but it also shows

---

particularly vulnerable to the impact of sleep deprivation.

[5] The U.S. Centers for Disease Control and Prevention has also labeled chronic partial sleep deprivation as a public health epidemic.

that the benefits of repeated group interaction may be entirely undone in the presence of SR group members.[6]

The literature on sleep and social interactions is somewhat limited. Studies that explicitly examine coordination games and SR are rare, notwithstanding the growing literature on sleep and decision making. McEvoy et al. (2021) report that SR participants are more likely to play Nash strategies over more complex strategies in simple 2-person coordination games. Others have examined how sleepiness (induced in various ways) reduces trust (Anderson and Dickinson, 2010; Dickinson and McElroy, 2017), dictator giving (Dickinson and McElroy, 2017; Ferrara et al., 2015), civic engagement (Holbein et al., 2019), or sensitivity to risk (Castillo et al., 2017; McKenna et al., 2007). Such results are consistent with related research showing that deliberative thinking, which is less likely with SR, is important for prosocial decisions (see also Chee and Chuah (2008); Krajbich et al. (2015); McCabe et al. (2001); Rilling and Sanfey (2011)).[7]

Because deliberation versus more automatic or intuitive decision process may help guide our understanding of how sleep restriction may impact decision making (e.g., Dickinson and McElroy (2019) explicitly note this and sketch a simple framework), the literature on dual process approaches to cooperation and prosociality is also relevant (e.g., Capraro (2019)). A meta-analysis recently reported that the dual process framework may not explain altruism that well (Fromell et al., 2020) unless explicitly considering gender differences. Specifically, intuition (as differentiated from deliberation) may promote altruistic behaviors in female but not male decision makers (Rand et al., 2016). Intuition has also been reported to increase cooperation that is non-strategic (i.e., more "pure" cooperation, see Rand (2016)), but a recent meta-analysis found rather weak evidence overall connecting intuitive thinking to cooperation unless the intuition was induced via emotions (Kvarven et al., 2020). Overall, the literature seems to suggest that many factors play a role in how relatively more versus less deliberative thinking may impact decision making in the more specific coordination setting that we study.

It may seem straightforward to hypothesize that sleep restriction will harm coordination

---

[6] Others have recognized how sleep may affect work team outcomes from an organizational standpoint (Barnes and Hollenbeck, 2009). Our experimental approach is intended to examine an environment where coordination success/failure can be clearly quantified.

[7] Another recent paper manipulated real world sleep in a noisy urban setting using a field experiment design to study how treatments to improve sleep may impact decision making and/or labor supply and productivity (Bessone et al., 2021). These authors reported no sleep level effect on decision making in their setting, which included estimated null effects on simple risk and 2-person social preferences. However, their estimated null effects must be interpreted carefully, because the approximate 30 minutes per night increase in sleep that resulted from their sleep treatments left participants with a level of sleep that would still be considered insufficient by public health experts (i.e., roughly 6 hrs/night). This highlights the challenges present in certain real world settings when trying to help people achieve well-rested sleep. Also, these authors did not explore coordination settings that are our focus. The value of field studies that use more real-world sleep levels/conditions to inform policy decisions has also been noted (Rao et al., 2021), but we believe coordination setting outcomes are worth more attention going forward.

outcomes because SR disproportionately impacts higher-level cognition and therefore probably harms most decision making.[8] However, coordination games contain key elements of risk and trust in the decision environment, and the literature suggests SR likely reduces trust but may increase willingness to take monetary risk. If successful coordination requires trust and/or willingness to take risk, then it is unclear what impact SR may have on coordination success/failure. Thus, it is premature or overly simplistic to assume SR will harm outcomes in coordination games. Additionally, the literature may not always carefully distinguish between whether an inferior outcome refers simply to coordination on a Pareto inferior effort choice or whether it refers to lower minimal effort within a group with others choosing higher effort (i.e., miscoordination costs due to wasted effort choices). It may not be straightforward to hypothesize how SR will impact either measure—a case can be made for hypothesizing either increased or decreased effort choices when SR as noted above, but then choice heterogeneity across SR individuals and/or within coordination groups comprised of both SR and WR individuals also seems an important empirical question to be answered from our data. This exploratory study provides first evidence on which of these SR effects likely dominates in a coordination setting.

3 Experimental Design

3.1 Sleep protocol

A large pool of potential participants were first administered validated instruments screened against the following exclusion criteria: extreme diurnal preferences (Adan and Almirall, 1991)[9], major depressive (Kroenke et al., 2003) and anxiety disorder (Spitzer et al., 2006), and diagnosed or suspected sleep disorders. Our study was also restricted to young adults between 18 and 40 years old. Viable participants who passed the screening criteria were randomly assigned, ex ante, to the well-rested (WR: 8-9 hr/night attempted sleep) or sleep-restricted (SR: 5-6 hr/night attempted sleep) treatment condition and then sent an invitations to participate in a one-week experiment that would involve the assigned nightly sleep level for 7 consecutive nights. Participants were informed they would be required to wear an actigraphy device to objectively measure sleep levels, keep a basic sleep diary provided by the experimenters, and participate in a 1.5 hour decision session at the end of the week.[10]

---

[8] As was noted above, a recent article examined EEG-based measures of working memory and concluded that this measure was positively related to successful coordination behavior (Mizrahi et al., 2020). Here, the authors examined 2-person coordination in a simple word choice task, but because sleep is known to affect working memory these results are relevant to our study.

[9] Excluding those with extreme morning or evening preferences, along with conducting all sessions between 10am-4pm on Tuesday-Thursday weekdays helped remove potential circadian or weekend sleep confounds in the data.

[10] The devices (specifically, the Actiwatch Spectrum Plus) have several advantages over lower cost commercial devices, and their validity for measuring sleep levels in non-disordered individuals is well-accepted

The study required two lab visits: Session 1 included informed consent procedures, survey instruments to collect data on a 6-item cognitive reflection task (Primi et al., 2015) and short-version of the Big Five personality measures (Gosling et al., 2003),[11] assignment of the actigraphy device and sleep diary, and participant QA. A cohort of (typically) 15-18 participants was recruited at a time (resulting in cohorts finishing the protocol with between 12-18 participants), and each cohort contained a mix of SR and WR participants to generate group heterogeneity in the coordination task. Sleep treatment assignments were private information and so participants were blinded to the sleep treatment condition of other participants in the cohort. It was explicitly noted (i.e., a common-knowledge announcement to the whole cohort by the experimenter) during Session 1 that sleep treatments were private information and others in the cohort may have different sleep treatment assignments.

Upon leaving Session 1, the experimenters emailed participants every 1-2 days to remind participants of their prescribed sleep levels, caution participants regarding sleepiness risks (to SR participants), and to remind all of the approaching decision Session 2. Because the at-home nature of the sleep protocol presents certain risks, it is important to note the risk management measures we employed (similar to those in Dickinson et al. (2017)). These measures included risk disclosure during informed consent procedures, regular cautionary emails during protocol week, and zero restrictions on compensatory behaviours like caffeine or sugar consumption. Because subjects were also free to withdraw from the study at any point and non-compliance to the sleep prescription did not produce large consequences (except possibly a reduction in their compensation in extreme cases), we (as experimenters) bore the extra cost burden of non-compliance in exchange for some additional risk mitigation in the at-home protocol.

Session 2 occurred one week after session 1 and included a short survey and self-report on sleepiness, decision task administration, and then the removal of actigraphy devices and cash payments based on decision experiment outcomes. Participants received a fixed compensation (check or Amazon gift code) of $25 for good-faith adherence to the prescribed sleep level, which was verified once actigraphy data were downloaded.[12]

We enrolled a total of n=128 treatment participants into the study (n=83 females; n=67 SR), but not all of those enrolled finished the protocol. In total, 101 participants (7 total cohorts of 12-18 participants per cohort) completed the one-week sleep treatment and decision experiment. Sleep watch data were corrupted for two participants, leaving complete sleep and decision data for 99 participants (n=62 females; n=46 SR). Sample selection was a

---

(Sadeh, 2011).

[11] We do not find significant pre-existing differences between compliant and not compliant participants in terms of CRT tests or personality ($p>.05$). The only significant pre-existing difference between SR and WR participants is regarding Openness (p-value$<.05$), which should not have an impact on coordination game decisions given anonymity within our decision task.

[12] Our standard for compliance with respect to the $25 payment was not as stringent as our standard for compliance for data analysis. In general, we chose to err on the side of paying all participants, with partial payments to the few who withdrew from the study during the treatment week.

concern due to study attrition. To address this important issue, we used observable characteristics from the preliminary recruitment sleep survey database to estimate the probability of study completion (see Appendix Table A1). Then, we constructed the inverse probability weights so that individual-level estimations reported below correct for this loss-to-follow-up attrition.[13]

3.2 Minimum effort coordination game

During the decision session, participants were administered the (incentivized) minimum effort coordination game through the Veconlab online platform.[14] The basic idea of the game is that members of a group must each decide on a level of hypothetical (but costly) effort. Once all decisions have been made, the payoff outcome in the game was dictated by the *minimum* effort choice within the group. Thus, lack of effort choice coordination implies wasted effort costs, and coordination at higher effort choices is payoff-preferred to coordination at lower effort choices.

For our study, the coordination game was played with groups of 3 members and the session administered a 10-round treatment using a Partner matching protocol and a 10-round treatment using a Stranger matching protocol (order of treatments counterbalanced across each cohort of participants).[15] Table 1 describes payoffs as a function of one's effort choice and the minimum effort choice of the other two players in one's group. Effort choice, e, has marginal cost of effort for each group member of $c = \$0.64$. Given this parameterization and our range of effort choice $e \in [1.1, 1.7]$ (choice option granularity was 0.01 units), each marginal increase of 0.1 effort units cost a member \$0.064. Among the set of Nash equilibrium outcomes there is a rank order of payoff preference such that maximal effort choice at e=1.7 for all group members is the payoff preferred Nash equilibrium. Our parameter choices were based on those in Goeree and Holt (2005) to approximate their "high-cost" of effort treatment that predicts minimal effort choices from among the set of Nash equilibrium predictions. Because each unit of effort chosen comes at a cost, it is worth noting that miscoordination can vary in severity.[16]

---

[13] Two cohorts of data are not analyzed here due to a change in our group size from 2-person to 3-person groups recruited for the majority of our sample (cohorts 3-9). However, we included all 9 cohorts of recruits (n=167) in the estimation of the selection model to improve efficiency of the inverse probability weights used in the data analysis of our 3-person groups.

[14] See http://veconlab.econ.virginia.edu/cg/cg.php for the Veconlab experiment page describing the coordination game.

[15] Because this design required a cohort or session of participants to be divisible by 3, we recruited a small number of backup participants for Session 2 who were not part of the cohort and who did not complete the sleep protocol. These backup participants helped ensure the use of all our treatment participant data in the event that our cohort size was not divisible by 3 due to participant attrition.

[16] Average effort choices shown in our Appendix Figure A1 can be compared to the high-cost treatment outcomes in Goeree and Holt (2005) where we see that convergence towards minimal effort is the norm in our data (understanding the exception in our data is what fuels our analysis and results shown in the next

Due to the multiplicity of Nash equilibria in the coordination game, an equilibrium refinement or selection criterion may help guide our baseline prediction. Given the parameterization we implement for our coordination games, the general prediction across a range of equilibrium selection criteria would be coordination on the Pareto-worst minimal effort level choice. This would be the case assuming risk dominance, applying a heuristic-based reference outcome criterion (Schneider and Leland, 2015)[17], or using a game-theoretic selection criterion that considers maximization of a potential function (Monderer and Shapley, 1996). For a useful example of how effort choice varies with cost of effort in this minimal effort choice environment, see also Goeree and Holt (2005). Based on existing literature on coordination games (see survey in Devetag and Ortmann (2007)), there are behavioural foundations to predict increased coordination at non-minimal effort levels when engaged in repeated interactions (i.e., the Partner matching protocol). Thus, established behavioural results may weaken the minimal effort prediction in our otherwise strong equilibrium selection prediction setting.

4 Results

4.1 Compliance, and manipulation check.

Sleep data on the 99 participants were scored using standard procedures and the objective nightly average sleep levels of subjects in the SR and WR treatment conditions are shown in Figure 1 (kernel density estimates). The main analysis utilized the full set of 101 participants who completed the protocol, and therefore considered treatment assignment as generating an "intent to treat" sample—this included two participants whose sleep watch data were corrupted given that the sleep data are not necessary for the dichotomous intent-to-treat scoring. We also considered the restricted sample of those deemed compliant with the prescribed sleep levels (slept < 375 min/night for SR, > 405 min/night for WR) based on scored actigraphy data. The (somewhat arbitrary) noncompliance region between 6.25 and 6.75 nightly hours of sleep (highlighted between the lines in Figure 1) is close to average nightly sleep levels in adults from recent survey evidence.[18] As such, the robustness analysis we conducted that removed these noncompliant participants can be thought of as removing

---

section).

[17] For example, the Reference Dependent Maximin criterion (Schneider and Leland, 2015) would predict minimal effect for all $c >$ \$0.50.

[18] See National Sleep Foundation. 2005. 2005 Sleep in America Poll. [Online] Available: http://www.sleepfoundation.org/sites/default/files/2005_summary_of_findings.pdf [accessed March 31, 2017]. See also more recent Gallup poll results (see http://www.gallup.com/poll/166553/less-recommended-amount-sleep.aspx [accessed March 31, 2017])

participants who were difficult to clearly classify as SR or WR in our data.[19]

Validity of the protocol is documented using data collected on self-report sleepiness using the validated Karolinska 9-point scale (Åkerstedt and Gillberg, 1990) as well a self-report of the extent to which the protocol altered one's typical sleep level ("self-report sleep gain/loss" range was [-4, +4] where 0 implied "no effect" on typical sleep levels). A fourth measure, Personal SD, was constructed to describe one's personal sleep deprivation level. This measure subtracted one's objective nightly sleep quantity from that participant's self reported nightly sleep-need for optimal performance, expressed in hours/night. Subjective sleep-need was elicited during the preliminary sleep survey at an earlier point in time and is therefore not endogenous with respect to one's treatment assignment. The Personal SD measure could be considered an individual-specific measure of one's level of SR (or WR) in our study, though bias in one's self-assessed sleep need is possible. For all measures and samples considered, we report a highly statistically significant difference between the SR and WR group ($p < 0.01$ in all instances; see 2, which reports results of Mann-Whitney nonparametric tests for differences in median values across treatment groups. It is clear that the sleep treatment we administered successfully manipulated objective and subjective sleep measures no matter whether considering the restricted sample or not.

The following sections organize our analysis of key outcome measures. Specifically, we first use regression analysis to explore determinants of key participant-level outcome measures. Here, the analysis relies on indicators and interactions between indicators variables for treatment condition, matching protocol, and round of play. Such regressions do, however, impose behavioral restrictions on the data. We therefore also examined the distributions of key individual outcome measures (*Effort* and *Earnings*) to further explore the impact of SR on different quantiles of these measures, and we also conducted stochastic dominance tests on the *Effort* and *Earnings* data. Then, we analyze key outcomes of the 3-person coordination groups in order to understand the impact of fewer versus more SR group members on the likelihood of equilibrium play and overall effort waste costs. Our analysis will show that SR group member *Effort* choices differ in important ways from those of WR group member *Effort* choices. And, at the group level, the evidence suggests a decreased likelihood of perfect coordination as well as increased cumulative wasted effort costs in groups with more SR members.

---

[19] Our compliance standard identifies more WR noncompliant subjects (n=14) than SR noncompliant subjects (n=2). In total, we have n=84 subjects deemed compliant. We should also note that, though we refer to "nightly" sleep levels in our participants, naps taken during the day were included in the sleep level calculations and in the continuous outcome measures of sleep used in the sensitivity analysis. Participants were encouraged to get their assigned level of sleep each night in one nightly (with no restrictions on exactly when during a night they go to bed and wake up), but if a participant took a nap they were encouraged to still note it in their sleep diary and those naps were identified in the actigraphy data record, scored, and added to the nightly sleep amounts. Thus, it was not the case that our "compliant" SR participants were likely not so compliant due to naps because naps were taken into account in what we call nightly sleep for ease of exposition.

4.2 Coordination game effort choices and earnings

We analyzed the following outcome measures from our full sample (intent-to-treat) data: effort choices, earnings, the likelihood of group coordination, and total effort waste costs. Table 3 reports results from models estimating individual Effort choice (column (1)), Earnings (column (2)), and the Gap between own effort and minimum effort in the group (column (3)) as a function of the treatment (Partner vs. Stranger matching protocol), treatment order (dummy variable for Partner condition in 2nd 10-round treatment), the Round ($= 1 - 20$), a dummy variable for assignment to the SR condition, and interactions between SR, Round, Partner matching, and Partner treatment order.[20] Similar models to those in Table 3 were estimated using only the subset of compliant participant data, model specifications that controlled for continuous *Personal SD*, and 2-stage instrumental variables models that used intent-to-treat SR assignment to predict Personal SD (SI Tables A2, A3).

Table 3 results in column (1) indicate a trend towards the minimal *Effort* choice prediction across rounds. Partner matching predicts significantly higher *Effort* when Partners occurs after the Strangers treatment, though marginally less so for SR participants. Our estimates indicate that SR participants choose somewhat higher effort levels than WR participants with Partner matching in general, though the precision on the *SR*Partner* interaction does not meet conventional significance levels. Other interaction terms are estimated to be statistically insignificant in column (1) of Table 3, though some trends seem apparent in the data. For example, effort choices in the final 5-rounds of a Partners treatment seem higher among SR compared to WR participants, though the regression analysis does not capture. We suspect this is due to the strong behavioral assumptions imposed by the interaction terms specification used in Table 3 (see SI Fig A1). This significant difference in late round effort choices can be highlighted in the cumulative distribution functions of *Effort* choices shown in Figure 2 (see Figure 3 for the corresponding late round *Earnings* distribution differences). The sensitivity analysis in Table A2 estimates this trend of increased effort in later rounds of the *Partners* treatment as statistically significant using an instrumental variables specification in models (5) and (6) (i.e., predicting one's *Personal SD* level by SR treatment assignment to reflect intent-to-treat).

We further document that SR increases *Effort* with a series of quantile regressions. As reported in the legend of Figure 2, the higher level of *Effort* by SR participants in late rounds occurs more broadly across quantiles of *Effort* choice in the Partners condition compared to the Strangers condition. Following the approach for consistent tests of stochastic dominance in Barrett and Donald (2003), we also conducted the full set of tests for first-order stochastic dominance (FOSD) of the late round *Effort* distributions in both the pooled data and separate treatment data. These tests utilized 50,000 bootstrapped resamplings to calculate p-values. In each of the samples (pooled, Strangers, Partners) the FOSD tests fail to reject

---

[20] To account for potential dependency of behavior across rounds (i.e., history of play), errors are clustered at the 3-person group level.

the hypothesis that the SR distribution FOSD the WR effort distribution (p>.10), while the opposite directional test *rejects* the hypothesis that the WR effort distribution FOSD the SR distribution (p<.10). In short, we find strong support for claiming the late rounds SR *Effort* distribution FOSD the WR distribution, and the dominance is more stark in the Partners condition as seen in Figure 2.

Column (2) of Table 3 shows results from estimating similar specifications with participant earnings, *Earnings*, as the dependent variable (see also SI Table A3). Here, we see that Partners matching generally increases *Earnings*, and *Earnings* also trend higher across rounds, which is likely due to a decrease in wasteful (uncoordinated) effort choices. There is no significant *SR*Partners* interaction in the *Earnings* regression. Trends in *average* earnings among SR participants, however, will not capture the fact that higher effort choices among SR participants may actually help increase *Earnings* of a group containing other SR individuals. As with our analysis of *Effort* choices, a look at the distribution of *Earnings* may be helpful here. The late rounds earnings distributions shown in Figure 3, which compares with late round *Effort* distributions shown in Figure 2, show evidence of increased variation in SR *Earnings* in later treatment, especially with Partners matching. There is at least some statistical support for this as well. The bottom of Figure 2 reports results from interquantile (IQ) regressions that document significant *Earnings* increases in most of the IQ regions for SR participants in the Partners treatment. We can also examine this by testing the full set of inequalities for second-order stochastic dominance (SOSD) similar to our FOSD tests conducted on effort choices (again, following Barrett and Donald (2003)). From these tests, we find at least marginal support that WR *Earnings* (late rounds) SOSD SR *Earnings* (fail to reject (p=.533) for the test that WR SOSD SR *Earnings*, but marginal rejection (p=.099) for the test that SR SOSD WR *Earnings*). The SOSD test combinations for the Strangers and Pooled data show failure to reject both sets of inequality tests (p>.10), from which we conclude no SOSD relationships.

Column (3) of Table 3 shows results from estimating similar specifications examining the gap between a participant's effort and the minimum effort of her group, *Gap*, which is our measure of the effort waste resulting from the participant's effort choice in each round.[21] The results partially reflect those of *Earnings* since this measures the distance to the group's earnings. We find that *Partners* and *Round* significantly reduce *Gap* but the *SR*Partners* interaction for *Gap* is not statistically significant at conventional levels.[22] The importance of wasted effort costs will be revisited in the group level data in Section 4.4 below.

---

[21] Note that our use of the term *Gap* may differ from its use elsewhere in the literature to refer to group-level miscoordination (e.g., see Feri et al. (2010)). In our paper, we are using *Gap* to refer to this individual-level measure of wasted effort, and we speak separately of effort-waste (miscoordination) costs experienced by the 3-person group as a whole.

[22] Analysis in the Appendix further examine the impact on SR and the gap between minimum and maximum *Effort* at the group level (see SI Tables A8 and A9) that lead to coordination failure costs.

4.3 Equilibrium play

At the group level, our interests turn to the examination of equilibrium play (i.e., successful coordination) and inefficiency costs of wasted effort choices. In general, our data show that groups populated with SR participants failed to coordinate significantly more often than groups populated with WR participants. Fisher's exact tests were used to document this general finding of decreased proportions of in-equilibrium play when the number of SR subjects in a group increases (see SI Appendix Table A4).

To more formally examine the likelihood of equilibrium play, Table 5 shows estimation results where the likelihood of coordination on any equilibrium outcome (no matter which of the multiple equilibria it is) was regressed on the number of SR participants in the 3-person group. We find that, even controlling for Round of play, additional SR group members significantly decrease the likelihood of effort coordination.[23] Table 5 also confirms the established result in the literature that Partners matching significantly increases the likelihood of successful coordination (Devetag and Ortmann, 2007) (in addition to increasing the level of effort to a more payoff-preferred equilibrium, as was noted in Table 3). Notably, the coefficient estimates on the SR dummy variables in Table 5 indicate that the increased miscoordination due to sleep restricted group members can be sufficiently high so as to negate the coordination-improving effect of Partner interactions.

4.4 Miscoordination costs

The previous analysis of equilibrium play likelihood does not consider that miscoordination (i.e., disequilibrium play) can vary in its severity. We define effort waste costs as the sum of group effort in excess of the minimum effort in the group. Figure 4 plots the cumulative distribution function of effort waste costs across all rounds of data both pooled and separated by treatment. Here, we see that cumulative group effort costs generally increase when a group contains SR members, and the introduction of just one SR member is sufficient to significantly increase effort waste costs. The difference is statistically significant comparing groups with zero versus all SR members (Kolmogorov test, $p<.001$) in tests on both Strangers and Partners treatment data and the result is robust to potential correlation between rounds of play.[24] We further test the distributional relationship of the pooled treatment data using the Barrett and Donald (2003) inequality tests for first-order stochastic dominance. Results here strongly support the following FOSD inequality relationships regarding the number

---

[23] Only in comparing 2-SR versus 1-SR subject groups in the Partners treatment do we find that an additional SR member increases the likelihood of coordination, although 2-SR member groups in Partners do not coordinate significantly more than 3-WR member groups (and significantly less than 3-SR member groups).

[24] In 10,000 bootstrap draws accounting for serial correlation in Partner sessions, we find that the Kolmogorov-Smirnov test is significant at the 5% level 93% percent of the time. Figure 4 reports the average p-value across bootstrap rounds. They are presented in parenthesis preceded by a "b".

of SR group members: 1SR $>$ 0SR, 2SR $>$ 0SR, 3SR $>$ 0SR, 2SR $>$ 1SR. These results also hold if using only the late rounds data. This key result documents the first evidence on how SR increases coordination failure and miscoordination waste, which are previously unidentified costs of insufficient sleep in cooperative dilemmas.

While it may seem intuitive to think that the costs of miscoordination are born largely by SR team members, recall that our previous analysis revealed a more complicated story. The cumulative distribution functions of late-round earnings in Figure 3 showed that SR individuals seemed to have a larger *variance* in earnings compared to WR individuals in the Partners condition, which suggests there were SR group members with both lower *and* higher earnings than WR group members in later rounds. Moreover, in operational settings a lack of successful coordination clearly imposes additional spillover costs to other stakeholders with an interest in team coordination.

5 Alternative Mechanisms Considered

As noted in Section 2, our data are informative regarding the net impact of SR on trust versus risk mechanisms in our setting. However, we can identify at least two other mechanisms for which there are testable implications from our data: gains from repeated play and Bayesian updating.

Regarding gains from repeated play, consider a two-effort level version of the game with payoffs to the *Low Effort* outcome=1 and to the *High Effort* outcome=2, and assume one's own effort costs $c < 1$ per unit. Table A5 summarizes the general-form game incentives=earnings are $1 - c$ if choosing *Low Effort*, and either $1 - 2c$ or $2 - 2c$ if choosing *High Effort* depending on the other's choice of low or high effort, respectively. Let p be the probability an opponent plays *High Effort*. We can define the basin of attraction of *Low Effort* as $p : 1 - c \geq (1 - p)(1 - 2c) + p(2 - 2c)$ or $p \in [0, c]$. Any time $p < c$, choosing *Low Effort* is more attractive. This is the basin of attraction of *Low Effort* for the strangers condition. In the Partners condition, it is possible to reduce the size of the *Low Effort* basin of attraction and therefore increase coordination at *High Effort*, which provides a theoretical underpinning to the superior coordination in Partners conditions result reported in the literature (Devetag and Ortmann, 2007). To simplify, consider that only two strategies are ever used: always choose *Low Effort* or choose *High Effort* until one player chooses *Low Effort* and then switch to *Low Effort* afterwards—these are analogous to *always defect* and *grim*, respectively, in repeated prisoners' dilemmas. Let q be the probability that an opponent uses *grim*. The basin of attraction of *Low Effort* is defined by the following equation: $q : (1 - c)T \geq (1 - q)(1 - 2c) + q(2 - 2c)T$. This leads to $q \in [0, \frac{[T-1-(T-1)c]+c}{2[(T-1)-(T-1)c]+1}]$. The upper bound of q is smaller than c for any $T > 1$ if $c > \frac{1}{2}$. While this strategy is defined for the first round of the game, it is still informative in latter rounds since it shows that the basin of attraction of *Low Effort* decreases in T. While this framework above does specify a model of behavior, it suggests that *Effort* is a function of the reward ($(a - c)e_t$), the maximum loss ($a\underline{e} - ce_t$), and the remaining number of rounds. Since we do not observe these variables, we may treat them as unobservables and test the prediction that *Round* affects choice differentially across in Partners versus Strangers.

A Bayesian framework assumes that both initial beliefs and new information impact one's belief regarding the likely payoff to a particular effort choice. We assume that the minimum group effort, *Min Effort*, in prior trials constitutes a type of new evidence, which is a noisier signal in the Strangers versus Partners condition that preserves the same group members: $evidence(Strangers) = evidence(Partners) + error$. It then follows that beliefs will be less sensitive to new evidence (i.e., prior trial *Min Effort*) in the Strangers condition. In our data, the hypotheses we test is that the previous trial *Min Effort* will positively impact current trial *Effort*, and this impact will be stronger in the Partners treatment compared to Strangers. Past research suggests SR reduces the weight one places on new information in a Bayesian setting (Dickinson and Drummond, 2008; Dickinson et al., 2016), and so we can also test whether the *SR*lag of Min Effort* interaction reduces this effect.

Table 4 presents estimation results to test these basic (alternative mechanisms) hypotheses.[25] As predicted by a basic Bayesian mechanism, the lag of *Min Effort* positively impacts current *Effort*. However, contrary to our predictions, this evidence weighting is statistically no different in the Partners compared to Strangers treatment, and also no different by SR assignment ($p = 0.263$). We observe that the estimated *Round* effect across the Partners and Strangers condition is not significantly different in either specification ($p = 0.688$), and thus it is not consistent with the "gains from repeated play" mechanism. However, SR participant *Effort* declines significantly slower in the Partner condition (*SR*Round* interaction). One might consider that SR promotes more best-response mistakes, but this would suggest (in our view) a difference in reinforcement learning that we should identify in our Bayesian evidence weight coefficient estimates test—we find no such differences. Of the mechanisms considered, this pattern in our data of a significantly slower effort choice decline among SR in the Partner condition is most consistent with a decreased sensitivity to the risk presented by higher effort choices made by SR group members. This result would be in line with the findings in Castillo et al. (2017) indicating decreased sensitivity to risk when sleepy.

6 Discussion and Conclusions

Our results and analysis are consistent with the hypothesis that successful coordination of choices decreases when more group members are sleep restricted, and this is most consistent with a decreased sensitivity to coordination risk when sleepy. This sheds light on previously unidentified costs of insufficient sleep levels in society and has implications for identifying

---

[25] In Strangers the previous round group is different for each subject. In Partners the previous round group is the same.

effective countermeasures. While our study quantifies some of these costs, in operational settings failed coordination also imposes spillover costs beyond the team that our study cannot quantitatively measure. Regarding countermeasures, wellness programs that focus on sleep health may be advisable, or there may be an increased need for hierarchy in an organization to improve coordination through supervisory control.

The impact of coordination failure due to sleepiness in field settings varies. For example, increased flight delays result when ground time depends on the minimal effort of the sleep-deprived team preparing an aircraft for its next flight. Or, sleepy community members (or local authorities) may hamper efforts to manage the spread of communicable disease, such as COVID 19, given that successful containment depends on the weakest link. One final example might be around-the-clock search and rescue efforts, which may face higher risk of failure or costly delays from miscoordination inefficiencies if team members suffer from insufficient sleep. These example also highlight that the inferior outcomes that result from coordination failure can have impacts that spread beyond the immediate group making the efforts. A main contribution of our study is that we combined elements of both field and laboratory methods in our examination of how a cognitive state impacts outcomes in an important cooperative dilemma environment. Specifically, the SR we induced is realistic and at levels commonly experienced in the "real world", the coordination game setting we used is controlled and generated quantifiable data, and our use of random assignment and validated objective sleep measurements facilitates our ability to claim SR has a causal impact on the behavioral outcomes measured.

The impacts of increasingly common insufficient sleep are somewhat well-known in terms of worker productivity, absenteesm, and adverse health effects—such impacts have been found to cost economies 1%-3% of their respective annual GDP (Hafner et al., 2017). In a sense, estimates sleepiness costs might conflate individual output declines and system-level loss of productivity due to miscoordination. A systematic study of how SR effects coordination efforts helps improve our overall understanding of how sleep impacts team outcomes, but such studies are lacking. We hope our study will help fill this void, draw attention to this area of inquiry, and stimulate interest in further examining how insufficient sleep impacts outcomes in other strategic settings.

Acknowledgements

DLD received support from Appalachian State University and the National Science Foundation under grants BCS-1229067 and SES-1734137. Many useful comments were provided by seminar participants at Monash University, the University of New South Wales, the University of Queensland, the University of Auckland, and the University of Montpellier.

References

Adan, A., Almirall, H., 1991. Horne and Ostberg morningness eveningness questionnaire - a reduced scale. Pers Indiv Differ 12, 241–253.

Åkerstedt, T., Gillberg, M., 1990. Subjective and objective sleepiness in the active individual. International Journal of Neuroscience 52, 29–37.

Anderson, C., Dickinson, D.L., 2010. Bargaining and trust: the effects of 36-h total sleep deprivation on socially interactive decisions. J Sleep Res 19, 54–63.

Barnes, C.M., Hollenbeck, J.R., 2009. Sleep deprivation and decision-making teams: Burning the midnight oil or playing with fire? Acad Manag Rev 34, 56–66.

Barrett, G.F., Donald, S.G., 2003. Consistent tests for stochastic dominance. Econometrica 71, 71–104.

Belloc, M., Bilancini, E., Boncinelli, L., DAlessandro, S., 2019. Intuition and deliberation in the stag hunt game. Scientific reports 9, 1–7.

Bessone, P., Rao, G., Schilbach, F., Schofield, H., Toma, M., 2021. The economic consequences of increasing sleep among the urban poor. The Quarterly Journal of Economics 136, 1887–1941.

Capraro, V., 2019. The dual-process approach to human sociality: A review. Available at SSRN 3409146 .

Castillo, M., Dickinson, D.L., Petrie, R., 2017. Sleepiness, choice consistency, and risk preferences. Theor Decis 82, 41–73.

Chee, M.W.L., Chuah, L.Y.M., 2008. Functional neuroimaging insights into how sleep and sleep deprivation affect memory and cognition. Curr Opin Neurol 21, 417–423.

Cooper, R.W., DeJong, D.V., Forsythe, R., Ross, T.W., 1990. Selection criteria in coordination games - some experimental results. Amer Econ Rev 80, 218–233.

Cooper, R.W., DeJong, D.V., Forsythe, R., Ross, T.W., 1992. Communication in coordination games. Q J Econ 107, 739–771.

Devetag, G., Ortmann, A., 2007. When and why? a critical survey on coordination failure in the laboratory. Exp Econ 10, 331–344.

Dickinson, D.L., Drummond, S.P., 2008. The effects of total sleep deprivation on bayesian updating. Judgment and Decision Making 3, 181.

Dickinson, D.L., Drummond, S.P., Dyche, J., 2016. Voluntary sleep choice and its effects on bayesian decisions. Behavioral Sleep Medicine 14, 501–513.

Dickinson, D.L., Drummond, S.P., McElroy, T., 2017. The viability of an ecologically valid chronic sleep restriction and circadian timing protocol: An examination of sample attrition, compliance, and effectiveness at impacting sleepiness and mood. PLOS ONE 12, e0174367.

Dickinson, D.L., McElroy, T., 2017. Sleep restriction and circadian effects on social decisions. Eur Econ Rev 97, 57–71.

Dickinson, D.L., McElroy, T., 2019. Bayesian versus heuristic-based choice under sleep restriction and suboptimal times of day. Games and Economic Behavior 115, 48–59.

Feri, F., Irlenbusch, B., Sutter, M., 2010. Efficiency gains from team-based coordinationlarge-scale experimental evidence. American Economic Review 100, 1892–1912.

Ferrara, M., Bottasso, A., Tempesta, D., Carrieri, M., de Gennaro, L., Ponti, G., 2015. Gender differences in sleep deprivation effects on risk and inequality aversion: evidence from an economic experiment. PLOS ONE 10.

Fromell, H., Nosenzo, D., Owens, T., 2020. Altruism, fast and slow? evidence from a meta-analysis and a new experiment. Experimental Economics 23, 979–1001.

Goeree, J.K., Holt, C.A., 2005. An experimental study of costly coordination. Game Econ Behav 51, 349–364.

Goldman, S.E., Stone, K.L., Ancoli-Israel, S., Blackwell, T., Ewing, S.K., Boudreau, R., Cauley, J.A., Hall, M., Matthews, K.A., Newman, A.B., 2007. Poor sleep is associated with poorer physical performance and greater functional limitations in older women. Sleep 30, 1317–1324.

Gosling, S.D., Rentfrow, P.J., Swann, W.B., 2003. A very brief measure of the big-five personality domains. J Res Pers 37, 504–528.

Hafner, M., Stepanek, M., Taylor, J., Troxel, W.M., Van Stolk, C., 2017. Why sleep mattersthe economic costs of insufficient sleep: a cross-country comparative analysis. Rand Health Quarterly 6, 11.

Holbein, J.B., Schafer, J.P., Dickinson, D.L., 2019. Insufficient sleep reduces voting and other prosocial behaviours. Nature human behaviour 3, 492–500.

Killgore, W.D.S., Grugle, N.L., Balkin, T.J., 2012. Gambling when sleep deprived: don't bet on stimulants. Chronobiol Int 29, 43–54.

Knez, M., Camerer, C., 2000. Increasing cooperation in prisoner's dilemmas by establishing a precedent of efficiency in coordination games. Organizational Behavior and Human Decision Processes 82, 194–216.

Krajbich, I., Bartling, B., Hare, T., Fehr, E., 2015. Rethinking fast and slow based on a critique of reaction-time reverse inference. Nat Commun 6.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2003. The patient health questionnaire-2: validity of a two-item depression screener. Medical care , 1284–1292.

Kvarven, A., Strømland, E., Wollbrant, C., Andersson, D., Johannesson, M., Tinghög, G., Västfjäll, D., Myrseth, K.O.R., 2020. The intuitive cooperation hypothesis revisited: a meta-analytic examination of effect size and between-study heterogeneity. Journal of the Economic Science Association 6, 26–42.

McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T., 2001. A functional imaging study of cooperation in two-person reciprocal exchange. P NATL ACAD SCI USA 98, 11832–11835.

McEvoy, D.M., Bruner, D.M., Dickinson, D.L., Drummond, S.P., 2021. Sleep restriction and strategy choice in cooperation and coordination games. Economics Letters 208, 110049.

McKenna, B.S., Dickinson, D.L., Orff, H.J., Drummond, S.P.A., 2007. The effects of one night of sleep deprivation on known-risk and ambiguous-risk decisions. J Sleep Res 16, 245–252.

Mizrahi, D., Laufer, I., Zuckerman, I., 2020. The effect of individual coordination ability on cognitive-load in tacit coordination games, in: NeuroIS retreat, Springer. pp. 244–252.

Monderer, D., Shapley, L., 1996. Potential games. Game Econ Behav 14, 124–143.

Ochs, J., 1995. Coordination problems. Handb Exp Econ , 195–252.

Pirrallo, R.G., Loomis, C.C., Levine, R., Woodson, B.T., 2012. The prevalence of sleep problems in emergency medical technicians. Sleep and Breathing 16, 149–162.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M.A., Hamilton, J., 2015. The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). J Behav Decis Making June 1.

Rand, D.G., 2016. Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. Psychological science 27, 1192–1206.

Rand, D.G., Brescoll, V.L., Everett, J.A., Capraro, V., Barcelo, H., 2016. Social heuristics and social roles: Intuition favors altruism for women but not for men. Journal of Experimental Psychology: General 145, 389.

Rao, G., Redline, S., Schilbach, F., Schofield, H., Toma, M., 2021. Informing sleep policy through field experiments. Science 374, 530–533.

Rilling, J.K., Sanfey, A.G., 2011. The neuroscience of social decision-making. Annu Rev Psychol 62, 23–48.

Sadeh, A., 2011. The role and validity of actigraphy in sleep medicine: an update. Sleep Med Rev 15, 259–267.

Schneider, M., Leland, J.W., 2015. Reference dependence, cooperation, and coordination in games. Judgm Decis Mak 10, 123–129.

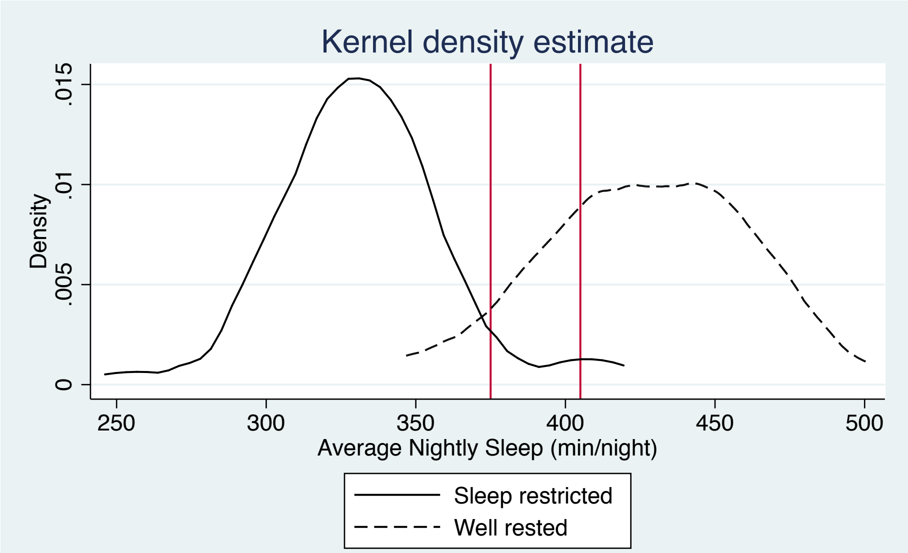
Schoenborn, C., Adams, P., 2010. Health behaviors of adults: United states, 2005-2007. Vital and Health Statistics. Series 10, Data from the National Health Survey , 1–132.

Spitzer, R.L., Kroenke, K., Williams, J.B., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the gad-7. Archives of internal medicine 166, 1092–1097.

Van Huyck, J.B., Battalio, R.C., Beil, R.O., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. The American Economic Review 80, 234–248.

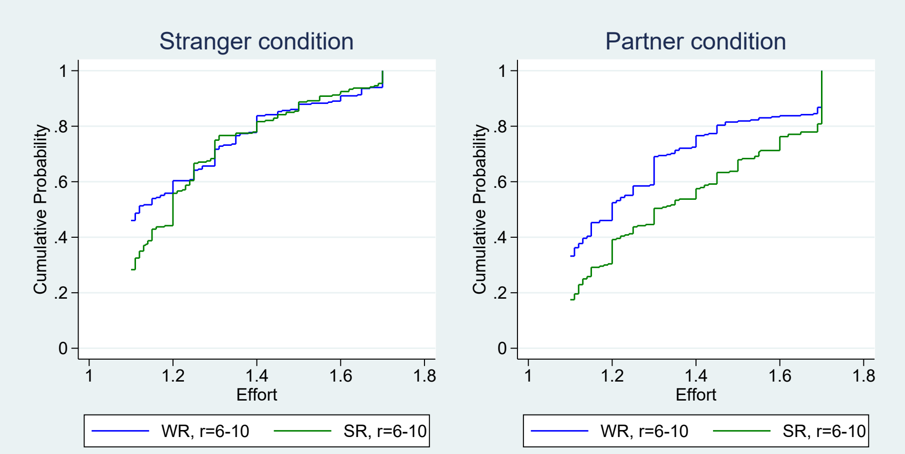
Van Huyck, J.B., Battalio, R.C., Beil, R.O., 1991. Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games. The Quarterly Journal of Economics 106, 885–910.

Figure 1: Treatment validation



(1) Nightly hours slept week prior to experiment (N=100), z=8.250, p-value<0.001
(2) Karolinska sleepiness scale (N=102), z=-5.939, p-value<0.001
(3) Self-reported sleep gain/loss (N=102), z=8.425, p-value<0.001
(4) Personal Sleep Deprivation (N=100), z=-5.611, p-value<0.001

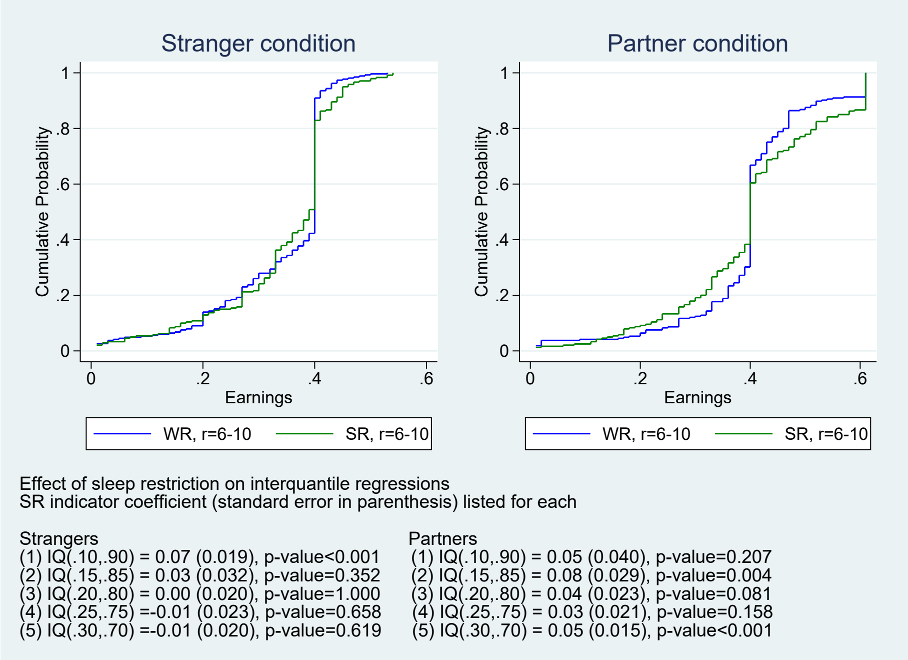
Figure 2: Effort Distributions (late rounds)



Effect of sleep restriction on quantile regressions
SR indicator coefficient (standard error in parenthesis) listed for each

Strangers	Partners
(1) Q(.15) = 0.00 (0.000), p-value= n.a.	(1) Q(.15) = 0.00 (0.003), p-value=1.000
(2) Q(.25) = 0.01 (0.003), p-value<0.001	(2) Q(.25) = 0.05 (0.012), p-value<0.001
(3) Q(.50) = 0.08 (0.019), p-value<0.001	(3) Q(.50) = 0.05 (0.028), p-value=0.075
(4) Q(.75) =-0.02 (0.025), p-value=0.431	(4) Q(.75) = 0.15 (0.053), p-value=0.005
(5) Q(.85) = 0.07 (0.036), p-value=0.051	(5) Q(.85) = 0.00 (0.026), p-value=1.000

Figure 3: Earnings Distributions (late rounds)

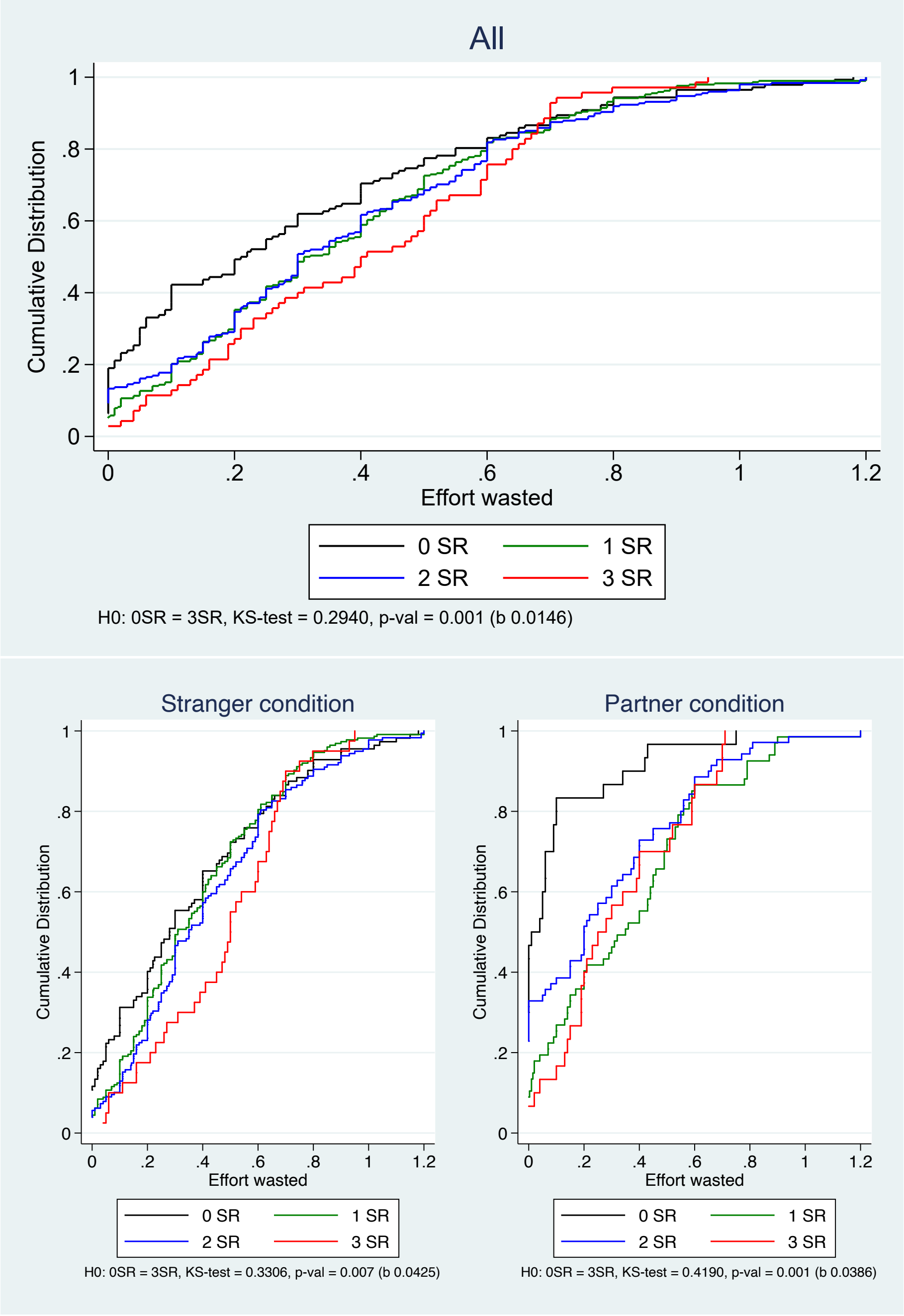


Effect of sleep restriction on interquantile regressions
SR indicator coefficient (standard error in parenthesis) listed for each

Strangers
(1) IQ(.10,.90) = 0.07 (0.019), p-value<0.001
(2) IQ(.15,.85) = 0.03 (0.032), p-value=0.352
(3) IQ(.20,.80) = 0.00 (0.020), p-value=1.000
(4) IQ(.25,.75) =-0.01 (0.023), p-value=0.658
(5) IQ(.30,.70) =-0.01 (0.020), p-value=0.619

Partners
(1) IQ(.10,.90) = 0.05 (0.040), p-value=0.207
(2) IQ(.15,.85) = 0.08 (0.029), p-value=0.004
(3) IQ(.20,.80) = 0.04 (0.023), p-value=0.081
(4) IQ(.25,.75) = 0.03 (0.021), p-value=0.158
(5) IQ(.30,.70) = 0.05 (0.015), p-value<0.001

Figure 4: Effort waste costs (all rounds)


All

H0: 0SR = 3SR, KS-test = 0.2940, p-val = 0.001 (b 0.0146)

Stranger condition

H0: 0SR = 3SR, KS-test = 0.3306, p-val = 0.007 (b 0.0425)

Partner condition

H0: 0SR = 3SR, KS-test = 0.4190, p-val = 0.001 (b 0.0386)

Table 1: Payoff matrix of coordination game

		Minimum of Other Members' Effort Choices						
		1.1	1.2	1.3	1.4	1.5	1.6	1.7
My effort choice	1.1	.396	.396	.396	.396	.396	.396	.396
	1.2	.332	.432	.432	.432	.432	.432	.432
	1.3	.268	.368	.368	.368	.368	.368	.368
	1.4	.204	.304	.404	.504	.504	.504	.504
	1.5	.140	.240	.340	.440	.540	.540	.540
	1.6	.076	.176	.276	.376	.476	.576	.576
	1.7	.012	.112	.212	.312	.412	.512	.612

Note: Light-shaded payoff cells reflect costly effort choice waste of other group members that do not directly impact one's own payoff.

Table 2: Protocol validation

Intent-to-treat	
Nightly hours slept week prior to experiment ($N = 100$)	$z = 8.250$, p-value< 0.001
Karolinska sleepiness scale ($N = 102$)	$z = -5.939$, p-value< 0.001
Self-reported sleep gain/loss ($N = 102$)	$z = 8.425$, p-value< 0.001
Personal Sleep Deprivation ($N = 100$)	$z = -5.611$, p-value< 0.001
Compliant subjects	
Nightly hours slept week prior to experiment ($N = 84$)	$z = 7.870$, p-value< 0.001
Karolinska sleepiness scale ($N = 84$)	$z = -5.336$, p-value< 0.001
Self-reported sleep gain/loss ($N = 84$)	$z = 7.748$, p-value< 0.001
Personal Sleep Deprivation ($N = 84$)	$z = -6.157$, p-value< 0.001

Note: Mann-Whitney tests. The number of observations for Intent-to-treat reflect two lost observations of sleep data from the actigraphy device measurements. Those two participants were still able to provide the self report measures of Karolinska sleepiness or sleep gain/loss

Table 3: Individual behavior

VARIABLES	(1) Effort	(2) Earning	(3) Gap
Partner cond. in last 10 rounds	0.116***	0.014	0.009
	[0.031]	[0.016]	[0.024]
Partner condition	0.030	0.044***	-0.051**
	[0.031]	[0.016]	[0.024]
Round	-0.007***	0.003***	-0.012***
	[0.002]	[0.001]	[0.002]
Sleep restricted	0.011	-0.014	0.037
	[0.032]	[0.018]	[0.025]
Sleep restricted × Partner cond.	0.053	-0.007	0.004
	[0.037]	[0.019]	[0.024]
Sleep restricted × Partner cond. last	-0.063*	-0.001	0.029
	[0.036]	[0.019]	[0.024]
Sleep restricted × Round	0.000	0.003	-0.007**
	[0.004]	[0.002]	[0.003]
Constant	1.315***	0.294***	0.395***
	[0.035]	[0.018]	[0.027]
Observations	2,020	2,020	2,020
R-squared	0.098	0.058	0.172

Robust standard errors in brackets, errors clustered at the 3-person group level. Models include correction for attrition using inverse probability weighting based on demographics and pre-experiment sleep survey information (see Appendix Table A1 for selection predictors)

*** p<0.01, ** p<0.05, * p<0.10

Table 4: Testing for risk dominance selection

VARIABLES	(1) Partners	(2) Strangers	(3) Partners	(4) Strangers
Min. effort (lagged)			0.782***	0.886***
			[0.044]	[0.069]
SR×Min. effort (lagged)			0.036	-0.157
			[0.067]	[0.101]
Round	-0.016***	-0.017***	-0.012***	-0.010***
	[0.004]	[0.003]	[0.003]	[0.004]
SR	-0.062	-0.015	-0.046	0.169
	[0.043]	[0.027]	[0.088]	[0.130]
Round×SR	0.018**	0.002	0.006	0.001
	[0.007]	[0.004]	[0.005]	[0.005]
Constant	1.414***	1.388***	0.426***	0.314***
	[0.030]	[0.022]	[0.060]	[0.093]
Observations	1,010	1,010	905	904
R-squared	0.027	0.052	0.420	0.258

Robust standard errors in brackets, errors clustered at the 3-person group level. Models include correction for attrition using inverse probability weighting based on demographics and pre-experiment sleep survey information (see Appendix Table A1 for selection predictors)

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table 5: Likelihood of perfect coordination

VARIABLES	(1) All	(2) Strangers	(3) Partners
Partner condition	0.069***		
	[0.021]		
Round	0.011***	0.007***	0.020***
	[0.002]	[0.002]	[0.005]
One SR subject in group	-0.080***	-0.038**	-0.166**
	[0.025]	[0.018]	[0.067]
Two SR subjects in group	-0.032	-0.048**	-0.035
	[0.027]	[0.022]	[0.068]
Three SR subjects in group$^+$	-0.061***	-	-0.106**
	[0.020]	-	[0.041]
χ^2 test on SR dummies	15.91	6.259	12.80
d.f.	3	2	3
p-value	0.001	0.044	0.005
1 SR = 2 SR	4.66	0.80	6.32
d.f.	1	1	1
p-value	0.0308	0.3721	0.0119
2 SR = 3 SR	0.80	-	6.75
d.f.	1	-	1
p-value	0.3721	-	0.0094
Number of clusters	401	365	38
Observations	760	340	380

Probit models (marginal effects). Robust standard errors in brackets, clustered at 3-person group level. $^+$ Dummy for 3-SR coordination group perfectly predicts failure of equilibrium play in Strangers.

*** p<0.01, ** p<0.05, * p<0.10

Online Supplementary Information Appendix

Human Subjects Protections

This study was reviewed and approved by the Office of Research Protections Institutional Review Board at Appalachian State University. The online preliminary survey was approved under IRB 09-0252, and the main study (sleep manipulation and decision making) was approved under IRB 16-0067. Consent for the online survey was obtained on p.1 of the survey (necessary for subjects to continue through the survey), and consent for the main study was obtained from participants at the beginning of Session 1.

Actigraphy Sleep Data Acquisition

Participants were assigned wrist-worn actigraphy devices commonly used in sleep research and validated against polysomnographic measures of total sleep time (Sadeh, 2011). Unlike commercial sleep trackers, these Actiwatch Spectrum Plus (Philips) devices provide subjects no sleep data feedback during the treatment week (this is only made known upon downloading the sleep data upon study completion) and the device batteries are sufficient to collect data the entire study week without recharge. Participants were also issued sleep diaries to complete daily and turn in at the end of the treatment week, and participants were required to send daily emails to the experimenter to report wake/bed times. The emails and sleep diaries were complementary to assist the actigraphy data scoring following procedures common to sleep studies (Goldman et al., 2007). The experimenter emailed participants about every 2 days to maintain contact, provide details and reminders of study parameters, caution participants regarding behaviours that might put them at risk if experiencing drowsiness as a result of participation in the experiment, and to remind them of the upcoming Session 2 that finalized the study.

Procedures

Enrollment into these experiments starts with an understanding of the database from which participants are identified and invited into the study. Each semester one of the authors (Dickinson) invites approximately 3000 students (mostly, a few under 40 years old faculty and staff perhaps) to complete the short online sleep survey that is used to collect basic demographic information in addition to the screening criteria measures. These regular surveys populate a database that he keeps updating every semester. These email lists are randomly generated lists of emails provided by central administration at his University. Of those 3000 emails received, it is typical for approximately 500 to complete the short survey, with the incentives being placement of ones name into a random drawing at the end of the academic year for a 95 USD Amazon gift code. They are also informed that the survey provides necessary information to become eligible for compensated sleep and decision making studies he runs. Then, of the 500 who complete the survey, not all meet the inclusion criteria for the study, and so the initial database-invitation group may result in perhaps 300 viable participants in a typical semester (those not at risk of major depressive or anxiety disorder, those without a diagnosed or suspected sleep disorder, no extreme diurnal preference, and within the 18-40 years old young adult age range). At this point, the viable participants are randomly assigned to a sleep condition (SR or WR) and then recruitment invitations are sent out. While a person randomly assigned to the SR condition could never then be recruited as a WR participant, it is possible that some invited for one study may not participate due to a scheduling or other conflict, but then they could be re-invited to participant with a future cohort where they may (or may not) enroll in the study due to session days/times better suited to his/her schedule that semester. When focusing on the selection issue regarding our experiments, we restrict our attention to what we feel is the more crucial question of whether one finishes the study or not, conditional on enrolling (not conditional on receiving an email invitation at some point, which we cannot even know whether it was seen by the participant).

The experiments were conducted in the APPEEL laboratory for experimental economics at Appalachian State University. Nine cohorts of participants were recruited, each containing a mix of participants who were randomly assigned the SR or WR sleep level in the study invitation email. Sleep treatment assignments were kept private in the laboratory session, and interactions in the coordination game were anonymous. All nine cohorts are used to for attrition analysis, but coordination game analysis is restricted to Cohorts 3-9 given our first two cohorts were administered the coordination game in 2-person groups, as opposed to the 3-person groups we used for the remaining cohorts (final samples sizes listed in paper are based on our cohorts 3-9). Decision sessions included three tasks in total, one of which was the coordination game. Participants received a fixed $25 compensation for compliance with the prescribed sleep levels (verified by actigraphy), and providing completed sleep diaries. Fixed compensation (by check or Amazon gift code) was paid several days after completion of Session 2, which was known to the participants, so that researchers could first download the sleep data and verify compliance efforts. Participants also received variable cash payoffs for outcomes in the decision experiments, including the coordination game. The coordination game task was computerized and administered through the Veconlab platform for experiments-the coordination game option used is at http://veconlab.econ.virginia.edu/cg/cg.php.

Table A1: Attrition - Probit on likelihood of completing study

VARIABLES	(1)
Female	0.055
	[0.266]
Minority	0.114
	[0.292]
Age	0.000
	[0.047]
Optimal Amount of sleep	-0.352**
	[0.145]
Sleep restricted treatment	-0.576**
	[0.243]
Depression Risk	0.382**
	[0.184]
Anxiety Risk	-0.083
	[0.053]
Constant	4.132**
	[1.634]
pseudo R2	0.020
Observations	167

Standard errors in brackets. Analysis includes subjects recruited for 2-player and 3-player sessions

*** p<0.01, ** p<0.05, * p<0.10

Table A2: Effort

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1		Model 2			
			OLS		IV	
VARIABLES	All	Compliant	All	Compliant	All	Compliant
Partner cond. in last 10 rounds	0.116***	0.125***	0.107***	0.139***	0.163***	0.157***
	[0.031]	[0.033]	[0.030]	[0.039]	[0.060]	[0.052]
Partner condition	0.030	0.023	0.006	0.008	-0.029	-0.025
	[0.031]	[0.032]	[0.030]	[0.038]	[0.059]	[0.052]
Round	-0.007***	-0.006***	-0.010***	-0.009***	-0.019***	-0.016***
	[0.002]	[0.002]	[0.002]	[0.003]	[0.006]	[0.005]
Sleep restricted	0.011	0.026				
	[0.032]	[0.031]				
Sleep restricted × Partner cond.	0.053	0.066*				
	[0.037]	[0.036]				
Sleep restricted × Partner cond. last	-0.063*	-0.058				
	[0.036]	[0.037]				
Sleep restricted × Round	0.000	-0.001				
	[0.004]	[0.004]				
Personal sleep deprivation			-0.015	-0.013	-0.062	-0.042
			[0.017]	[0.021]	[0.042]	[0.036]
Per. sleep depriv. × Partner cond.			0.027**	0.026	0.042	0.040*
			[0.013]	[0.016]	[0.027]	[0.022]
Per. sleep depriv. × Partner cond. last			-0.011	-0.024	-0.041	-0.035
			[0.013]	[0.016]	[0.027]	[0.022]
Per. sleep depriv. × Round			0.001	0.001	0.006**	0.005*
			[0.001]	[0.001]	[0.003]	[0.003]
Constant	1.315***	1.292***	1.351***	1.330***	1.444***	1.394***
	[0.035]	[0.037]	[0.040]	[0.051]	[0.088]	[0.078]
Observations	2,020	1,660	1,980	1,660	1,980	1,660
R-squared	0.098	0.106	0.103	0.105	0.065	0.083

Errors clustered at the 3-person group level. Models include correction for attrition using inverse probability weighting based on demographics and pre-experiment sleep survey information (see Appendix Table A1 for selection predictors).
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table A3: Earnings

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1		Model 2			
			OLS		IV	
VARIABLES	All	Compliant	All	Compliant	All	Compliant
Partner cond. in last 10 rounds	0.014	0.013	-0.010	-0.005	0.015	0.012
	[0.016]	[0.017]	[0.016]	[0.021]	[0.032]	[0.026]
Partner condition	0.044***	0.050***	0.055***	0.064***	0.057*	0.065***
	[0.016]	[0.016]	[0.016]	[0.020]	[0.031]	[0.025]
Round	0.003***	0.003***	0.006***	0.007***	0.010***	0.010***
	[0.001]	[0.001]	[0.001]	[0.002]	[0.003]	[0.003]
Sleep restricted	-0.014	-0.015				
	[0.018]	[0.018]				
Sleep restricted × Partner cond.	-0.007	-0.013				
	[0.019]	[0.020]				
Sleep restricted × Partner cond. last	-0.001	0.002				
	[0.019]	[0.020]				
Sleep restricted × Round	0.003	0.003				
	[0.002]	[0.002]				
Personal sleep deprivation			0.018*	0.021*	0.038*	0.036**
			[0.009]	[0.011]	[0.022]	[0.018]
Per. sleep depriv. × Partner cond.			-0.007	-0.010	-0.007	-0.009
			[0.007]	[0.008]	[0.014]	[0.011]
Per. sleep depriv. × Partner cond. last			0.012*	0.010	-0.001	0.001
			[0.007]	[0.008]	[0.015]	[0.012]
Per. sleep depriv. × Round			-0.001**	-0.002**	-0.003**	-0.003**
			[0.001]	[0.001]	[0.001]	[0.001]
Constant	0.294***	0.293***	0.256***	0.247***	0.218***	0.217***
	[0.018]	[0.018]	[0.022]	[0.027]	[0.045]	[0.038]
Observations	2,020	1,660	1,980	1,660	1,980	1,660
R-squared	0.058	0.063	0.069	0.074	0.047	0.059

Errors clustered at the 3-person group level. Models include correction for attrition using inverse probability weighting based on demographics and pre-experiment sleep survey information (see Appendix Table A1 for selection predictors).
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table A4: Equilibrium play

All

	SR subjects in group				
	0	1	2	3	Total
Off equilibrium	115	281	217	68	681
Percent	0.81	0.94	0.87	0.97	0.90
In equilibrium	27	17	33	2	79
Percent	0.19	0.06	0.13	0.03	0.10

Fisher's exact test p-value < 0.001

Partners

	SR subjects in group				
	0	1	2	3	Total
Off equilibrium	52	132	109	28	321
Percent	0.74	0.94	0.78	0.93	0.84
In equilibrium	18	8	31	2	59
Percent	0.26	0.06	0.22	0.07	0.16

Fisher's exact test p-value < 0.001

Strangers

	SR subjects in group				
	0	1	2	3	Total
Off equilibrium	63	149	108	40	360
Percent	0.88	0.94	0.98	1.00	0.95
In equilibrium	9	9	2	0	20
Percent	0.12	0.06	0.02	0.00	0.05

Fisher's exact test p-value $= 0.009$

Table A5: Two-effort level coordination game

	1	2
1	1-c, 1-c	1-c,1-2c
2	1-2c,1-c	2-2c,2-2c

Table A6: Distance to group minimum effort

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1		Model 2			
			OLS		IV	
VARIABLES	All	Compliant	All	Compliant	All	Compliant
Partner cond. in last 10 rounds	0.027**	0.031**	0.047***	0.055***	0.041	0.043*
	[0.012]	[0.014]	[0.015]	[0.019]	[0.027]	[0.023]
Partner condition	-0.033***	-0.042***	-0.053***	-0.062***	-0.068**	-0.074***
	[0.012]	[0.013]	[0.015]	[0.019]	[0.027]	[0.022]
Round	-0.006***	-0.005***	-0.010***	-0.010***	-0.017***	-0.016***
	[0.001]	[0.001]	[0.001]	[0.002]	[0.003]	[0.003]
Sleep restricted	0.017	0.023				
	[0.018]	[0.019]				
Sleep restricted × Partner cond.	0.027	0.037*				
	[0.018]	[0.020]				
Sleep restricted × Partner cond. last	-0.020	-0.022				
	[0.018]	[0.020]				
Sleep restricted × Round	-0.003	-0.004*				
	[0.002]	[0.002]				
Personal sleep deprivation			-0.024**	-0.026**	-0.060***	-0.051***
			[0.010]	[0.012]	[0.021]	[0.019]
Per. sleep depriv. × Partner cond.			0.017**	0.019**	0.022	0.023*
			[0.008]	[0.009]	[0.014]	[0.012]
Per. sleep depriv. × Partner cond. last			-0.015**	-0.018**	-0.013	-0.013
			[0.008]	[0.009]	[0.014]	[0.012]
Per. sleep depriv. × Round			0.002***	0.002**	0.006***	0.005***
			[0.001]	[0.001]	[0.001]	[0.001]
Constant	0.181***	0.173***	0.231***	0.233***	0.302***	0.286***
	[0.013]	[0.015]	[0.020]	[0.025]	[0.039]	[0.034]
Observations	2,020	1,660	1,980	1,660	1,980	1,660
R-squared	0.053	0.054	0.064	0.064	0.036	0.041

Errors clustered at the 3-person group level. Models include correction for attrition using inverse probability weighting based on demographics and pre-experiment sleep survey information (see Appendix Table A1 for selection predictors).
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table A7: Distance to group median effort

	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1		Model 2			
			OLS		IV	
VARIABLES	All	Compliant	All	Compliant	All	Compliant
	[0.002]	[0.002]				
Partner cond. in last 10 rounds	0.001	0.006	0.010	0.007	-0.017	-0.002
	[0.009]	[0.010]	[0.012]	[0.013]	[0.023]	[0.017]
Partner condition	-0.015	-0.029***	-0.018	-0.025*	-0.017	-0.040**
	[0.009]	[0.009]	[0.012]	[0.013]	[0.022]	[0.017]
Round	-0.004***	-0.004***	-0.005***	-0.005***	-0.009***	-0.009***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]
Sleep restricted	0.015	0.014				
	[0.014]	[0.015]				
Sleep restricted × Partner cond.	-0.005	0.010				
	[0.014]	[0.014]				
Sleep restricted × Partner cond. last	0.017	0.014				
	[0.014]	[0.014]				
Sleep restricted × Round	-0.002	-0.002				
	[0.002]	[0.002]				
Personal sleep deprivation			-0.003	-0.004	-0.026	-0.026*
			[0.006]	[0.008]	[0.018]	[0.014]
Per. sleep depriv. × Partner cond.			0.000	0.001	-0.002	0.007
			[0.005]	[0.006]	[0.012]	[0.009]
Per. sleep depriv. × Partner cond. last			0.000	0.004	0.014	0.008
			[0.005]	[0.006]	[0.012]	[0.009]
Per. sleep depriv. × Round			0.000	0.000	0.003**	0.003**
			[0.000]	[0.000]	[0.001]	[0.001]
Constant	0.131***	0.130***	0.141***	0.142***	0.185***	0.187***
	[0.009]	[0.010]	[0.015]	[0.017]	[0.034]	[0.028]
Observations	2,020	1,660	1,980	1,660	1,980	1,660
R-squared	0.045	0.047	0.040	0.043	0.016	0.025

Errors clustered at the 3-person group level. Models include correction for attrition using inverse probability weighting based on demographics and pre-experiment sleep survey information (see Appendix Table A1 for selection predictors).
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table A8: Gap at the group level

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) IV	(5) OLS	(6) IV
Partner condition	-0.051***	-0.053***	-0.050***	-0.053***	-0.052***	-0.057***
	[0.018]	[0.018]	[0.018]	[0.020]	[0.018]	[0.022]
Round	-0.012***	-0.014***	-0.012***	-0.012***	-0.015***	-0.020
	[0.001]	[0.003]	[0.001]	[0.002]	[0.003]	[0.012]
No. of sleep restricted subjects in group	0.023**	0.010				
	[0.010]	[0.017]				
No. of SR subjects in group × Round		0.001				
		[0.002]				
Mean Personal Sleep Deprivation of subjects in group			0.010	0.090**	-0.006	0.045
			[0.011]	[0.044]	[0.019]	[0.078]
Mean Pers. SD of subjects in group × Round					0.002	0.004
					[0.002]	[0.007]
Constant	0.381***	0.399***	0.394***	0.255***	0.423***	0.335**
	[0.021]	[0.031]	[0.026]	[0.079]	[0.037]	[0.144]
Observations	760	760	760	760	760	760
R-squared	0.167	0.168	0.158	0.051	0.159	0.051

Robust s.e. in brackets, errors clustered at the group level
*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table A9: Group Dynamics

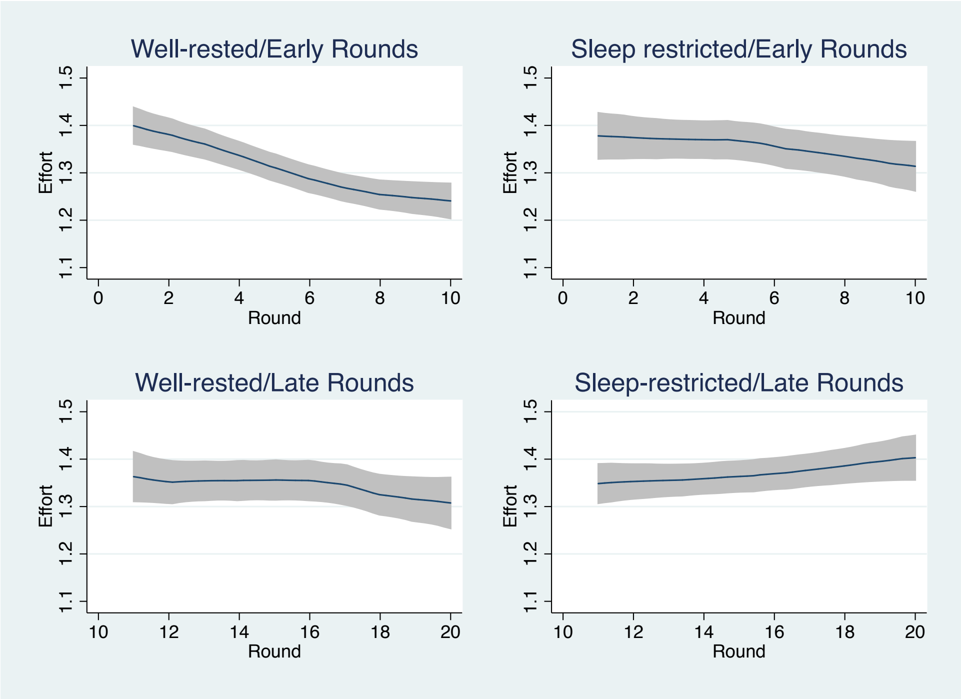
	OLS						IV		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Min	Max	Gap	Min	Max	Gap	Min	Max	Gap
Partner condition	0.077***	0.026	-0.051***	0.076***	0.025	-0.050***	0.076***	0.024	-0.053***
	[0.023]	[0.025]	[0.018]	[0.022]	[0.023]	[0.018]	[0.023]	[0.024]	[0.020]
Round	-0.002	-0.014***	-0.012***	-0.002	-0.014***	-0.012***	-0.002	-0.014***	-0.012***
	[0.002]	[0.002]	[0.001]	[0.002]	[0.002]	[0.001]	[0.002]	[0.002]	[0.002]
No. of SR subjects in group	0.007	0.029**	0.023**						
	[0.015]	[0.014]	[0.010]						
Mean Personal SD				0.045***	0.055***	0.010	0.026	0.114**	0.089**
				[0.013]	[0.013]	[0.011]	[0.056]	[0.051]	[0.044]
Constant	1.171***	1.552***	0.381***	1.100***	1.495***	0.394***	1.135***	1.391***	0.257***
	[0.027]	[0.029]	[0.021]	[0.029]	[0.033]	[0.026]	[0.095]	[0.092]	[0.079]
Observations	760	760	760	760	760	760	760	760	760
R-squared	0.064	0.161	0.167	0.113	0.188	0.158	0.104	0.139	0.054

Robust s.e. in brackets, errors clustered at the group level

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Figure A1: Effort by round and experimental condition

Partners



Strangers