Circadian Rhythms and Physiological Synchrony: Evidence of the Impact of Diversity on Small Group Creativity

EUNICE JUN, University of Washington, USA
DANIEL MCDUFF, Microsoft Research, USA
MARY CZERWINSKI, Microsoft Research, USA

Circadian rhythms determine daily sleep cycles, mood, and cognition. Depending on an individual's circadian preference, or chronotype (i.e.,"early birds" and "night owls"), the rhythms shift earlier or later in the day. Early birds experience circadian arousal peaks earlier in the morning than night owls. Prior work has shown that individuals are more effective at analytic tasks during their peak arousal times but are more creative during their off-peak times. We investigate if these findings hold true for small groups. We find that time of day and a group's majority chronotype impact performance on analytic and creative tasks. Physiological synchrony among group members positively predicts group satisfaction. Specifically, homogeneous groups perform worse on all tasks regardless of time of day, but they achieve greater physiological synchrony and feel more satisfied as a group. Based on these findings, we present and advocate for a temporal dimension of group diversity.

CCS Concepts: • **Human-centered computing → User interface toolkits**; **Empirical studies in collaborative and social computing**; *Social engineering (social sciences)*;

Additional Key Words and Phrases: circadian rhythms; physiology; creativity; cognition; group work; group creativity; diversity; productivity

ACM Reference Format:
Eunice Jun, Daniel McDuff, and Mary Czerwinski. 2019. Circadian Rhythms and Physiological Synchrony: Evidence of the Impact of Diversity on Small Group Creativity. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 60 (November 2019), 22 pages. https://doi.org/10.1145/3359162

1 INTRODUCTION

Intelligent systems for group work and classroom learning abound. These systems rely on sensing various aspects of group interaction, ranging from tracking spatial movement [12] to measuring voice and social interactions using sociometric badges [43]. Using this data, systems can provide feedback on group behavior and health. More recently, researchers have begun to use physiological signals to gauge engagement [31] and increase awareness of partner task load [53]. However, no system has yet considered the most pervasive physiological pattern: the circadian rhythm. To understand how the circadian rhythm might be leveraged in intelligent systems, an understanding of its impact on groups is first necessary.

Circadian rhythms have widespread influence on human functioning, including sleep [13], mood [10, 38–40], and cognition [34, 45]. Individuals have circadian preferences, called chronotypes.

Authors' addresses: Eunice Jun, University of Washington, Seattle, Washington, 98195, USA, emjun@cs.washington.edu; Daniel McDuff, Microsoft Research, Redmond, Washington, 98052, USA, damcduff@microsoft.com; Mary Czerwinski, Microsoft Research, Redmond, Washington, 98052, USA.

2573-0142/2019/11-ART60 $15.00
https://doi.org/10.1145/3359162

Some individuals are "early birds" who experience the diurnal patterns of peaks and valleys in circadian arousal earlier in the day. Morning chronotypes often awaken, prefer to start working, and feel tired earlier in the day. Others are "night owls" who experience the patterns later in the day and tend to awaken later, stay up working, and go to bed later.

Chronotypes are more than habitual preferences. Chronotypes also determine the times of day individuals are more focused or distractible. The phases of the circadian cycle where circadian arousal is high (peaks) are characterized by increased on-task focus [28], and people exhibit increased analytical abilities and memory at these times [32, 34, 64]. On the other hand, the off-peak times of circadian arousal are characterized by increased distraction but also increased creative problem solving ability [32, 64].

Despite growing evidence that circadian rhythms impact individual cognition [34, 64] and productivity [26, 27], the impact of these rhythms on group work has been overlooked. Furthermore, recent work in human-computer interaction has found that physiological synchrony impacts group satisfaction and group performance on a wide variety of tasks related to collective intelligence [67]. However, circadian rhythms, the driver of many physiological processes, has yet to be considered in the work on the physiology of group work. With the availability of unobstrusive wearable sensors, it is possible to study these phenomena.

To fill the gaps in knowledge identified above, we investigate the impact of circadian rhythms on group performance on analytic and creative tasks, group physiological synchrony, and group satisfaction. Studying the impact of circadian rhythms on group work is important not only because small groups and organizations face creative and strategic challenges but also because the power of groups may lie in their resilience to individual susceptibility to circadian rhythms.

Our overarching research question is "How do circadian rhythms affect group creativity and collaboration?". Based on an in-lab study with 28 groups of two or three people, we find that groups are not immune to the influence of circadian rhythms. The majority's circadian peaks determine when the group will be most effective at solving analytic and creative problems. Furthermore, groups where all members have the same chronotype achieve higher levels of physiological synchrony, but they also perform worse than diverse groups at all times of day.

Our experimental results have important theoretical and practical applications for human-computer interaction. Specifically, we

(1) Show that group performance shifts throughout the day due to group composition, extending prior work on the impact of circadian rhythms on creativity and cognition in individuals to the small group context.
(2) Provide evidence that there is a physiological basis for why people prefer working in groups who are similar to them even though diversity improves group outcomes.
(3) Define implications for productivity, group work scheduling, experimental practices, and diversity.

2 BACKGROUND

Our interdisciplinary work builds upon methods and findings in physiology, psychology, and human-computer interaction.

2.1 Circadian Rhythms and Cognition

The impact of circadian rhythms in our daily lives is pervasive. In animals and human beings, circadian rhythms govern various physiological processes that maintain homeostasis, including sleep, digestion, and thermoregulation [37]. In humans, the circadian timing system affects how we adapt to environmental and temporal changes [15, 51], such as shift work and productivity [26, 27].

The circadian cycle is consistent across human beings. There is a global peak shortly after awakening, a second, smaller peak about halfway through the waking hours, and then a steady decline over the course of the day. These shifts correspond to levels of cortisol and melatonin, which are thought to chemically regulate the wakefulness-sleep cycle [13]. Individuals exhibit shifts in their circadian cycles, which constitute their chronotype [29], or circadian preference. In some individuals, the onset of the diurnal pattern is earlier than others. As a result, these individuals have morning chronotypes ("early birds") and tend to awaken, start work, and sleep earlier than those whose cycles start later (evening chronotypes, or "night owls").

Medical researchers more recently have found that genetic (i.e., melanopsin genotype) and environmental factors, such as the duration of daylight (e.g., seasonality) affect circadian rhythms [48]. In the presence of the TT P10L melanopsin genotype, participants tended towards increased morningness on shorter days. On shorter days, people may generally tend to wake up closer to sunrise to conduct business or maximize exposure to sunlight, for instance. Nonetheless, people who have morning chronotypes on longer days will likely still tend towards morningness more than people with evening chronotypes, even during shorter days.

Circadian cycles and chronotypes also impact human cognition. Research has found that the peaks of circadian arousal correspond to increased cognitive inhibition, which impacts information processing abilities [28]. Hasher, Zacks, and May [28] posit that cognitive inhibition plays three key roles in information processing: access, deletion, and restraint. At peak times, *access* to working memory is limited to only task-relevant thoughts and stimuli, thoughts or ideas that were helpful but no longer are suitable for the strategy at hand are *deleted*, and incorrect solutions and strategies are *restrained*. The effect of increased inhibition is the dedication of working memory to the task and the current strategy. Inhibitory control diminishes with age and at off-peak hours [34].

The theory of inhibition explains the results of various empirical studies where researchers find that participants make more analytic decisions [9] and remember and recall more accurately [34, 45] during their circadian arousal peaks. However, reduced cognitive inhibition during off-peak times leads to broader exploration of details and strategies needed to solve creative puzzles where seemingly irrelevant details are key to the solution [33, 64].

More recently, researchers have found further evidence supporting the theory of inhibition. Delpouve et al. found that individuals can accurately estimate their optimal and nonoptimal times of day based on their chronotype scores [21]. During times of day that individuals defined to be their non-optimal, the researchers found increased implicit learning abilities, as measured by an artificial grammar learning task, suggesting that a reduction in subjective alertness increases individuals' abilities to think more broadly [21].

There appear to be cognitive benefits to both chronotypes. Compared to evening chronotypes, morning chronotypes are able to allocate cognitive resources more efficiently, suggesting increased cognitive adaptability, during suboptimal times [42]. On the other hand, evening chronotypes may be more resilient to sleep deprivation [7]. The mechanisms and implications of each chronotype's strengths are still under investigation. Our work contributes broadly to recent work in this area by providing a group-based perspective.

2.2 Insight and Incremental Problem Solving

To test the theory of inhibition and the effect of circadian patterns on cognition, researchers have used two kinds of problems: insight and incremental problems. Insight problem solving abilities are measures of convergent creativity [20, 22]. Insight problems require solvers to think of novel and useful solutions [22], many times involving "lateral thinking." Incremental problems encompass problems such as algebra problems that require applying a rule or strategy to linearly reach a solution.

There are important cognitive differences between insight and incremental problems that make them apt for use to study the impact of chronotype and time of day. While solving insight problems, people usually arrive at an impasse that requires them to challenge their assumptions and restructure a problem space [17, 50]. Solutions usually arrive with an "Aha!" moment that is ineffable [49], difficult to predict [35], and subject to different hindsight biases than incremental problems [6, 16]. In incremental or analytic problems, on the other hand, the constraints and process to a correct answer are usually straightforward and do not require re-representation [49], making prediction of how close a solver is to a solution more accurate [35].

In our work, we focus on convergent creativity because small groups and organizations face challenges where one creative, strategic decision or solution must be applied even though many options may be considered along the way. We operationalize convergent creativity by using insight problems and follow the precedent set in the psychological literature.

Using classic insight and incremental problems, Wieth and Zacks [64] found that participants were significantly more likely to answer insight problems more accurately, and by extension, be more effective creative thinkers, during their non-optimal times of day. There was a trend but no statistical significance in the opposite direction for accuracy on incremental problems.

The category of problems that researchers have found to elicit insight has expanded to include not only classic insight problems but also compound remote associate tasks and anagrams [61]. Researchers argue that contextualizing questions with participants' subjective ratings of the level of insight required to accurately answer a problem is important in this classification [60]. However, in the group-based setting, it is unclear how these individual ratings translate to a group's collectively perceived levels of insight necessary to solve a problem. Instead of eliciting group ratings, we follow best practice guidelines and utilize multiple insight problems that appeal to verbal, spatial, and mathematical skills and do not rely on intelligence tests and noninsight problems as counterparts to insight problems [61]. Instead, we pilot and carefully curate our insight and incremental problems to be ones that are likely to appeal to different cognitive processes.

Like Wieth and Zacks, the vast majority of the research on the impact of circadian rhythms on creative problem solving has focused on the individual. We differentiate our work from and extend this research on time of day effects on cognition by examining small groups. We also augment the work by drawing on recent work in HCI that investigated the physiological characteristics and structures of collaboration.

2.3 The Physiology of Collaboration

Group synchrony is a critical part of many important social experiences, such as marches, music festivals, and religious rituals. Group synchrony supports group cohesion and cooperation even when particular individuals may have to sacrifice for the group [66]. Synchrony increases group prosocial behavior [58], affiliation [30], and rapport [57]. There are two distinct forms of synchrony: behavioral and physiological. Behavioral synchrony, such as when dancers dance in complete unison together, is key to group coordination. More difficult to detect, physiological synchrony in electrodermal activity (EDA), heart rate (HR), and facial expression also plays an important role in group social bonding.

Monster et al. [36] monitored participants' EDA and electromyography while groups of three folded origami boats. The authors found synchrony in movements of the zygomaticus major muscle (smiling) and EDA. Although synchrony in smiling was positively related to feelings of team cohesion, synchrony in skin conductance predicted feelings of otherness and negative affect for the other members.

Furthermore, Chikersal et al. [14] found that synchrony in EDA mediated the positive impact of gender and ethnic diversity on dyads' group satisfaction. Synchrony in facial expressions mediated the negative impact of age difference and the positive impact of ethnic diversity on collective intelligence, a general measure of a groups' aptitude that can predict future group success [67].

The impact of physiological synchrony on group cohesion, group satisfaction, and collective intelligence are clear. However, there is no unified understanding of why physiological synchrony may predict some negative group social outcomes even though researchers have found that synchrony repeatedly leads to greater social cohesion.

We posit that the varied impact of synchrony is due to the impact of group composition on a physiological level–namely, circadian patterns and chronotypes. Because circadian rhythms drive physiological responses at the individual level, we theorize that circadian rhythms are also key in group cohesion. To test and build upon this theory, we study problem solving during circadian peak and off-peak times and focus on synchrony in EDA and heart rate, two physiological processes that are more immune to conscious control than facial expression.

2.4 Hypothesis Development

Based on prior work, we developed specific hypotheses for three broad categories: cognitive performance, physiological synchrony, and group satisfaction.

2.4.1 Cognitive Performance. Based on previous findings we expect that groups will exhibit increased performance on creative tasks during their off-peak times and have improved performance on incremental tasks during their peak times. We use the group's majority chronotype to determine the group's peak and off-peak circadian times of day. Cognitive performance can be quantified as accuracy and time spent on tasks.

H1: Groups will answer incremental problems more accurately during the majority's peak time than during their off-peak time.
H2: Groups will answer insight problems more accurately during the majority's off-peak time than during their peak time.

We hypothesize that homogeneous groups, where all members are of the same chronotype, and heterogeneous groups will perform differently. We hypothesize:

H3: Homogeneous groups will be more accurate on incremental problems during their peak-times than heterogeneous groups, who have the same majority chronotype. Similarly, homogeneous groups will be more accurate on insight problems during their off-peak times than heterogeneous groups.

2.4.2 Physiological Synchrony. Regardless of chronotype, people exhibit predictably similar diurnal patterns of circadian arousal. Chronotypes determine how phase shifted these patterns are. People of the same chronotype experience the phases at the same times, but people of different chronotypes are at different, and sometimes opposed, phases throughout the day. Therefore, we expect that homogeneous and heterogeneous groups will be able to achieve physiological synchrony to different degrees when working together:

H4: Homogeneous groups will experience greater degrees of physiological synchrony than heterogeneous groups.

2.4.3 Group Satisfaction. Because peaks in circadian rhythms correspond to peaks in positive affect, we expect that people will enjoy working in groups more during their peak times. As a result, we hypothesize that group chronotype and the time of day will impact group satisfaction.

H5: Groups who work in the majority's peak time will report higher group satisfaction.

Researchers have previously found that physiological synchrony positively predicts group cohesion [36, 66], affiliation [30], and rapport [57]. Therefore, we hypothesize that:

H6: Groups that attain greater physiological synchrony, regardless of group chronotype composition, will experience increased group satisfaction.

3 EXPERIMENTAL DESIGN

We conducted a laboratory experiment modeled most closely after [14] and [64] to test our hypotheses. We focused on small groups of two and three because prior work on time of day effects on creativity and cognition has only been conducted with individuals [33, 64] and because HCI research with groups has thus far focused on physiology and collaboration with dyads [14]. This study received IRB approval.

Our experiment had four independent variables: group majority *chronotype* (morning vs. evening), group *composition* (homogeneous vs. heterogeneous), the *time of day* groups participated (morning vs. late afternoon), and *question type* (insight vs. incremental). Group chronotype, composition, and time of day were between subjects variables, and problem type was a within subjects variable.

4 STUDY

4.1 Participants

We recruited 40 participants (mean age=33.62 years, std age=8.7 years) from a large technology company. 25 participants (10 female, 15 male) had morning chronotypes, and 15 (5 female, 10 male) had evening chronotypes. Participants were screened to have morning or evening chronotypes at the time of recruitment using the Morningness-Eveningness Questionnaire [29]. We describe this process in further detail below in subsection 4.3. Participants represented a range of occupations, including software engineer, project manager, and researcher. All participants were compensated $100 for their time.

4.2 Procedure

Participants came into the lab for two sessions on the same day (one morning and one afternoon). During both experimental sessions, they completed pre-session surveys, completed two sets of problems (one alone, another in a group of two or three) while wearing a wrist-worn sensor, and then completed post-surveys.

The morning session was from 8:15 to 9:30am and the afternoon from 4:30pm to 5:30pm. These times were chosen based on prior research protocols [64] and fit into normal workday hours. At the beginning of the first morning session, participants gave informed consent and completed a pre-study survey that assessed their present affective state and demographics on a laptop, wore an E4 [24] sensor on their nondominant wrist, and then began laptop cameras for video and audio recording.

Afterwards, participants completed six problems (three insight and three incremental). Participants had four minutes to complete each problem (24 minutes total). All problems were completed on paper and a timer was shown on laptops in front of each participant. Each problem was presented

Table 1. **Group composition.** Individuals in the experiment participated in different groups in the morning and afternoon. Groups consisted of either two or three individuals who were either heterogeneous or homogeneous in chronotype. From left column to right: Individual **participants** that comprise the group and their chronotypes (blue = evening, orange = morning); the **group's majority chronotype**, and **if the time of day (of experimental testing) was optimal/nonoptimal** given the group's majority chronotype.

Participants	**Group Chronotype**	**Optimal?**
Time of Day: **Morning**		
Heterogeneous groups		
20 28 34	Evening	Nonoptimal
04 18 27	Morning	Optimal
38 40 43	Morning	Optimal
42 45 46	Morning	Optimal
07 17 21	Morning	Optimal
09 16 35	Morning	Optimal
Homogeneous groups		
10 25 33	Morning	Optimal
11 12 22	Morning	Optimal
47 48 49	Morning	Optimal
15 23 24	Morning	Optimal
32 36	Morning	Optimal
03 26 29	Evening	Nonoptimal
05 30	Evening	Nonoptimal
Time of Day: **Afternoon**		
Heterogeneous groups		
01 22 33	Morning	Nonoptimal
02 11 25	Morning	Nonoptimal
10 12 19	Morning	Nonoptimal
42 43 49	Evening	Optimal
03 05 23	Evening	Optimal
15 26 30	Evening	Optimal
18 28	Morning*	Nonoptimal
24 29	Evening*	Optimal
Homogeneous groups		
21 35 36	Morning	Nonoptimal
38 46 48	Morning	Nonoptimal
17 32	Morning	Nonoptimal
40 47	Morning	Nonoptimal
27 34	Morning	Nonoptimal
07 09	Evening	Optimal
04 20	Evening	Optimal

one at a time at the top of a sheet of paper, which participants could use to work out their solutions. At the end of the four minutes, participants circled their answers and then moved on to the next question. If participants completed a problem before the four minutes, they wrote the remaining time (as shown on the timer) at the top of the problem page. After the sixth question, participants completed a post-session survey that asked if participants had seen the problems before and, if participants worked in a group, and asked questions about group satisfaction.

After an optional restroom break, participants completed their second morning session with a new block of six questions and a second post-session survey.

The protocol sans the informed consent and pre-study survey was repeated for the afternoon sessions.

To control for learning and ordering effects, we counterbalanced the order of questions, and group/individual sessions. All participants solved all 24 problems (see Appendix A) and worked in a group in the morning and with another, different group in the afternoon. Participants never worked with the same person in a group twice.

Participants worked in groups of two (nine groups) or three (19 groups). Participants were originally assigned to groups of three but some worked in groups of two because of no-shows. There were a total of 14 homogeneous and 14 heterogeneous groups. Ten homogeneous groups were formed with participants who all had morning chronotypes, and four homogeneous groups were formed with only evening chronotypes. The majority of eight heterogeneous groups had morning chronotypes, four groups had an evening chronotype majority, and two dyads were evenly split (one morning and one evening). An overview of participant chronotype and group composition during the morning and afternoon sessions are given in Table 1.

4.3 Measures

To test our hypotheses, we were interested in group composition, cognitive performance, physiological synchrony, and group satisfaction. In this section, we define each experimental variable and discuss its operationalization in our experiment.

Group composition. We were interested in forming homogeneous and heterogeneous groups based on individuals' chronotypes. To assess chronotype, we used the validated Morningness-Eveningness (MEQ) questionnaire [29]. The MEQ asks 19 questions about physical performance at different times of day, situational actions to manage sleep debt, and sleep preferences. We used the MEQ as a screening instrument to recruit the intended number of homogeneous and heterogeneous groups each day. Individuals can score into five categories of chronotype: definite morning, moderate morning, neutral, moderate evening, and definite evening. Consistent with prior research [33, 64, 68], we did not recruit participants who scored in the neutral category and combined the definite and moderate groups to form two chronotype categories of interest: morning and evening.

Cognitive performance. To assess group analytic ability, we use incremental problems [49]. To measure creative ability, we use insight problems [20, 22]. We assess accuracy as well as the mean time (in seconds) taken to arrive at correct answers. The highest accuracy score for the incremental and insight problems was three, and the longest time duration for a single question was four minutes.

To develop our battery of incremental and insight problems, we aggregated problems from prior research and publications [3, 63, 64]. We then tested these problems in a pilot and then eliminated and replaced any questions that were too confusing or too easy, as determined by time taken to solve them. Prior to administering the problems, we developed an answer book that we used to score the solutions.

Participants either received one point for a correct response or zero for an incorrect response. There was no partial credit. Because all questions were open-ended and could be answered with drawings, paragraphs, and matrices, two authors scored a random subset of the answers by strictly applying the answer book and discussing discrepancies before coming to consensus. One author then graded the remaining questions. Inter-rater reliability was found to be quite satisfactory (κ = .88).

We include the 24 problems used in the experiment in the appendix.

Physiological synchrony. We assessed physiological synchrony throughout the entire 24 minutes by measuring electrodermal activity (EDA) and heart rate (HR) using the wrist-worn E4 sensor [24].

EDA and HR are measures of arousal that can be used to assess stress, engagement, and excitement within individuals and organizations [4, 41, 44]. EDA and HR differ, eccrine sweat glands are entirely under sympathetic control [11, 18] and therefore is a pure measure of arousal. HR measures arousal in the sympathetic and parasympathetic nervous systems and is correlated with [18], but not the same as EDA. To obtain a more holistic view of arousal, we measure both.

EDA signals have two components: phasic and tonic. The phasic EDA component captures the local changes (ragged ups and downs) of the signal, and the tonic EDA component indicates the coarser patterns (rising and falling) of the EDA signal. HR was derived from the raw BVP (Blood Volume Pulse) data recorded on the E4.

Heart Rate: A 6th-order Butterworth filter (cut-off frequencies of 0.7 and 3 Hz) was applied to the blood volume pulse (BVP) signals measured from the E4 (at a rate of 64Hz) to smooth the signal. A 5s moving window (with step size 5s) was then used to successfully calculate the heart rate of the individual throughout the 24 minute period. For each 5s second segment, a fast fourier transform (FFT) was applied and the highest peak within the range 0.7 to 3 Hz (42 to 180 BPM) was selected as the heart rate.

EDA: The E4 recorded EDA data at a rate of 4Hz. We translated the time series data into signals over a period of 24 minutes that participants worked together. To extract the phasic and tonic components of the EDA we used the NeuroKit[1] toolbox. The α and γ parameters were set to $8x10^{-4}$, 10^{-2} respectively. The skin conductance response (SCR) method used was "makowski" and the threshold applied was 0.1.

Since there might be some individual differences in the timing of physiological responses to the stimuli we applied Dynamic Time Warping [8]. For the DTW we allowed a maximum warping of 5 seconds (i.e., the maximum time difference between samples from the respective time series was 5s). With these signals, we then calculated synchrony across the entire 24 minutes of group work using the Pearson correlation coefficient. In cases where there were three group members, the correlation of the group as a whole was calculated as the mean of the correlations between the pairwise combinations.

Figure 2 shows examples of the heart rate and EDA signals for two groups working in the afternoon: a homogeneous group with a morning chronotype (off-peak) and a heterogeneous group with a majority evening chronotype (peak). Notice how the homogeneous group's EDA and HR are highly correlated while the heterogeneous groups' EDA and HR are not.

Group Satisfaction. We selected and adapted 10 questions from the validated Team Diagnostic Survey (TDS) [59]. Our questions came from two higher-order constructs: team interpersonal processes and individual learning and well-being. We included six straightforward and four reverse questions. Participants responded to statements such as "Working together energizes and uplifts members of our team." on a five-point scale (1 ="highly inaccurate" and 5 = "highly accurate.")

[1]https://github.com/neuropsychology/NeuroKit.py

Table 2. Accuracy Model. * denotes significance

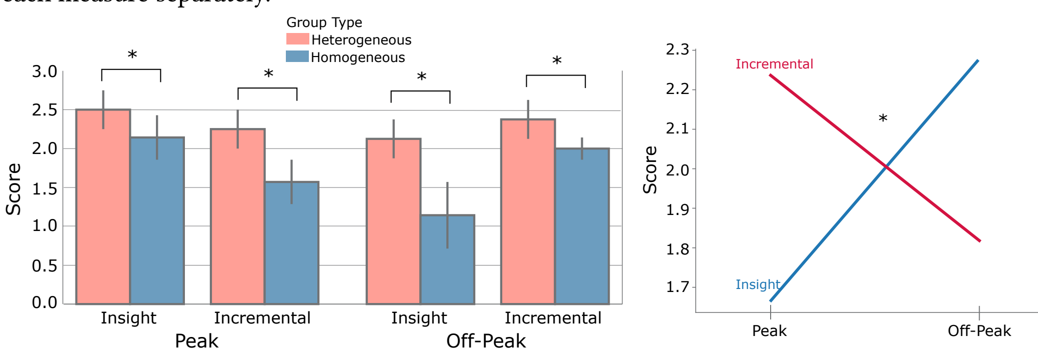
Model	Coeff(SE)	t	d
Constant	2.02(.11)	18.96	<.0001*
Question Type	.02(.11)	.19	.85
Time of Day	.11(.11)	1.01	.32
Group Composition	-.31(.11)	-2.88	.006*
Time of Day x Ques. Type	-.25(.11)	-2.33	.024*

Because group satisfaction is a group-level construct, we calculate a group's mean and use the means in our analyses.

5 RESULTS

5.1 Cognitive Performance

To test for performance, we considered accuracy and task time to arrive at an accurate solution for each measure separately.



(a) Group composition effect (b) Circadian peak/off-peak effects

Fig. 1. Group Results on Problems. (a) Heterogeneous groups outperform homogeneous groups at all times of day on both insight and incremental tasks. (b) Groups are more accurate at solving incremental than insight problems during their peak time, the opposite is true during the off-peak time. There is a significant interaction effect between time of day and question type.

Accuracy: To test our first hypothesis that groups will have increased accuracy on incremental problems during the majority's peak time and increased accuracy on insight problems during the majority's off-peak time, we built a least squares linear regression model with the predicted variable as the score (number of questions each group answered correctly) and the factors as question type (QT) (incremental vs. insight), time of day (TOD) (peak vs. off-peak), the interaction between question type and time of day, and group composition (GC) (homogeneous vs. heterogeneous).

$$Score = QT + TOD + QT * TOD + GC \tag{1}$$

The overall model was statistically significant ($F(4,47)=3.57$, $p=.013$, $R^2=.23$). Table 2 summarizes the model. Group composition had a main effect on accuracy ($F(1,47)=8.28$, $p<.01$). We expected homogeneous groups to outperform heterogeneous groups at incremental tasks during their peak time and at insight tasks during their off-peak time (H3). However, we found that homogeneous groups always had lower accuracy than heterogeneous groups (see Table 2).

Table 3. Time to Correct Answer Model. * denotes significance

Model	Coeff(SE)	t	d
Constant	128.55(7.95)	16.17	<.0001*
Question Type	36.62(7.95)	4.61	<.0001*
Time of Day	-1.68(8.09)	-.21	.84
Group Composition	2.32(8.01)	.29	.77
Time of Day x Ques.Type	-3.91(7.96)	-.49	.63

The interaction between question type and time of day also had a statistically significant effect on group accuracy (($F(1,47)=5.43$, $p<.05$). As predicted (H1,H2), groups performed better on incremental tasks during the majority's peak circadian times and better on insight tasks during the majority's off-peak times. When groups answered incremental tasks during the majority's off-peak hours, they performed worse than they did during their peak-hours, as seen in Figure 1 and Table 2.

Time to correct answer: To test our second hypothesis that groups would be faster to accurately answer incremental problems during their peak time and faster to accurately answer insight problems during their off-peak time, we developed a least squares linear regression model. We included time (in seconds) to arrive at a correct solution as the predicted variable and question type (incremental vs. insight), time of day (peak vs. off-peak), the interaction between question type and time of day, and group composition (homogeneous vs. heterogeneous) as the factors. All samples in which the question was not answered correctly were excluded from the model.

$$Time = QT + TOD + QT * TOD + GC \tag{2}$$

The overall model was statistically significant ($F(4,44)=5.74$, $p=.0008$, $R^2=.34$). Table 3 summarizes the results. Question type was the only significant effect ($F(1,44)=21.24$, $p<.0001$). Controlling for time of day, the interaction between question type and time of day, and group composition, we found that groups took significantly more time to correctly answer incremental problems than to solve insight problems (see Table 3).

Group composition: Using the accuracy model above, we tested our hypothesis that homogeneity would have a synergistic effect on groups' accuracy. We expected homogeneous groups to be better at incremental tasks during their peak time and to be better at insight tasks during their off-peak time, compared to mixed groups who had the same majority chronotype but also included someone from the opposite chronotype. Conversely, we expected homogeneous groups to be worse at insight tasks during their optimal time and worse at incremental tasks during their non-optimal time.

Controlling for the question type, time of day, and the interaction, homogeneous groups always had lower accuracy than heterogeneous groups ($F(1,50)=8.28$, $p<.01$). Contrary to our hypothesis, heterogeneous groups are always more accurate throughout the day (see Figure 1a).

5.2 Physiological Synchrony

Group: Due to similarities in circadian arousal patterns, we expected homogeneous groups to experience greater degrees of physiological synchrony than heterogeneous groups. To examine physiological synchrony, we calculated correlations in phasic EDA, tonic EDA, and HR among group members across group compositions. Using t-tests between homogeneous and heterogeneous groups, we compared the signed Pearson correlations as well as the absolute values of the correlations due to a concern that directionality differences might wipe out important relationships in the

Table 4. Physiological synchrony between homogeneous and heterogeneous groups. Homogeneous groups reach greater degrees of physiological synchrony in phasic EDA. Absolute correlations capture the presence of a relationship between physiological measures irrespecrtive of direction. Physiological measures that are in the opposite direction indicate there may be some compensation or polarization affect, which a directional correlation would not capture and we would like to measure.* denotes p<.05

	Homo. μ	**Hetero.** μ	**t-value**
phasic EDA, dir.	.18	-.01	-2.08*
phasic EDA, abs.	.24	.18	-.87
tonic EDA, dir.	.12	.11	-.10
tonic EDA, abs.	.19	.16	-.44
HR, dir.	-.02	-.10	-.55
HR, abs.	.34	.24	-1.26

group data. Comparing the absolute correlations helps us isolate the degree from the directionality of the relationship between members in homogeneous and heterogeneous groups.

Using a one-sided t-test, we found that homogeneous groups achieved greater synchrony in directional phasic EDA (t(26)=-2.08, p<.05). Table 4 details all the comparisons between both groups.

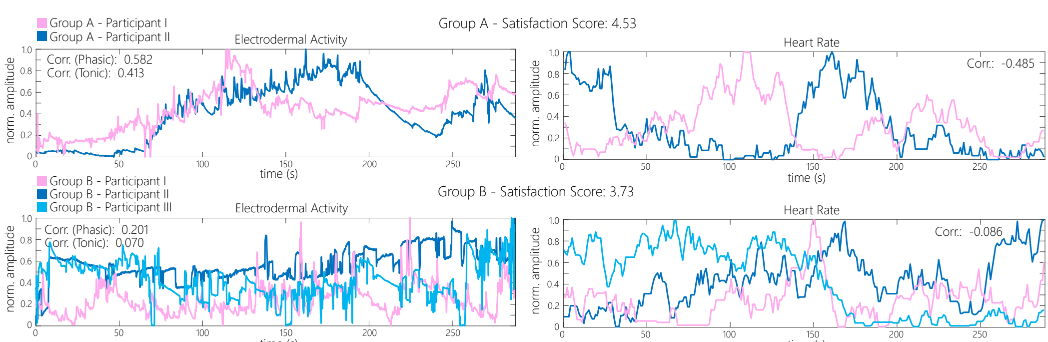


Fig. 2. Physiological synchrony examples. (a) Group A, a homogeneous group of two, had high group satisfaction (4.53), high synchrony in EDA (phasic=.58, tonic=.41), and high synchrony in HR (-.48). (b) Group B, a heterogeneous group of three, had lower group satisfaction (3.73) and low synchrony in EDA (phasic=-.20, tonic=-.07) and HR(-.09). Overall, homogeneous groups experienced greater synchrony in phasic EDA.

5.3 Group Satisfaction

To test our hypotheses that increased physiological synchrony and peak times of day would positively predict group satisfaction, we built a least squares linear regression model, which included group satisfaction as the predicted variable and the three (phasic and tonic EDA and HR) absolute physiological correlations, group composition (homogeneous vs. heterogeneous), and time of day as the factors. We chose to include the absolute correlations because we were interested in relationships, regardless of directionality.

$$Satis. = EDA_{phasic,r} + EDA_{tonic,r} + HR_r + GC + TOD \tag{3}$$

The model was statistically significant (F(5,17)=2.90, p<.05). Table 5 summarizes the results of the model.

Table 5. Group Satisfaction Linear Model. * denotes significance

Model	Coeff(SE)	t	d
Constant	3.66(.15)	24.90	<.0001 *
Phasic EDA(abs.)	-.08(.36)	-.23	.82
Tonic EDA(abs.)	1.12(.44)	2.52	.02 *
Heart Rate(abs.)	.77(.33)	2.32	.03 *
Time of Day	-.13(.06)	-2.22	.04 *
Group Composition	.08(.06)	1.34	.20

Time of day (F(1,17)=4.93,p=.04), tonic EDA synchrony (F(1,17)=6.35,p=.02), and HR synchrony (F(1,17)=5.37, p=.03) were statistically significant main effects.

Time of day: Our hypothesis that groups working in the majority's peak time would report higher group satisfaction was confirmed. We found that controlling for physiological synchrony and group composition, groups that worked during their off-peak time of day reported lower group satisfaction, as shown in Table 5.

Physiological Synchrony: Our hypothesis that physiological synchrony positively affects group satisfaction was confirmed. We found that controlling for group composition and time of day, synchrony in tonic EDA and HR positively predicted group satisfaction, as seen in Table 5. Synchrony in phasic EDA did not significantly determine group satisfaction.

Finally, to analyze how physiological synchrony might affect performance, we conducted a stepwise regression, where we added all physiological measures, group composition, question type, time of day, and the interaction between question type and time of day to predict accuracy and time. We found that the best models for predicting performance were still our original models for score and time without the physiological measures. Table 6 gives an overview of our results and the corresponding hypotheses.

5.4 Group interactions

Groups were instructed to solve the problems together. The research team observed and made notes on group behavior and problem solving strategies.

Organization and decision making strategies varied between groups. For most problems and groups, individuals read a problem and each offered a strategy for solving the problem, a possible solution, and/or support for or disagreement on previously explicated ideas. In some instances across incremental and insight problems, the research team observed that an individual would drive the problem solving, explain to the rest of the team, and try to convince the others that their solution was correct. No group operated solved problems solely in one way or another.

6 DISCUSSION

We found that group chronotype and time of day affected group performance on analytic and creative tasks. Groups were more effective at analytical tasks during the majority's circadian peak but were more effective at creative tasks during the majority's off-peak time. We also found that physiological synchrony did not affect group performance. However, physiological synchrony in heart rate predicted how satisfied groups were working together. Although heterogeneous groups were more effective at problem solving at all times of day, homogeneous groups reached greater degrees of physiological synchrony.

Homogeneous groups achieved higher levels of synchrony in phasic EDA. Physiological synchrony is associated with increased feelings of affiliation [30] and trust [57]. Our findings suggest

Table 6. Summary of significant effects. The rows represent three outcome measures we wanted to explain. indicate positive findings and - null findings. If we had hypotheses for a particular variable predicting an outcome measure, the are enumerated in parentheses. All findings on physiolgoical synchrony differences between groups (H4) can be found in Table 4. GC = Group Composition, TOD = Time of Day, QT = Question Type, PS = Physiological Synchrony

	GC	TOD	QT	QT x TOD	PS
Accuracy	$(H3)$	-	-	$(H1, H2)$	-
Time	-	-		$(H1, H2)$	-
Satisfaction	-	$(H5)$	-	-	$(H6)$

that chronotype similarity can be predictors of these same forms of group bonding. Synchrony in tonic EDA and heart rate predicted group satisfaction. Unlike [14], we did not find that synchrony in phasic EDA affected group satisfaction. Furthermore, we found the opposite of [36]. We found that synchrony in EDA positively predicted feelings of group satisfaction, not tension. This suggests that synchrony in tonic and phasic EDA may be susceptible to more fundamental physiological processes, such as circadian rhythms, are important for different kinds of tasks, and/or are dependent on group composition.

Our physiological synchrony and group satisfaction results provide physiological evidence as to why people may prefer to work with others who are similar to them. Homogeneous groups experienced greater synchrony, which, based on prior research, would predict enhanced social cohesion. However, homogeneous groups always performed significantly worse than heterogeneous groups. Therefore, even though working with people of similar physiological predispositions may be more enjoyable, working with physiologically diverse people can have significant creative and analytic benefits.

Finally, our findings that groups have optimal times of day for analytic and creative tasks that depend on members' chronotypes is noteworthy. The timing of group work could determine the productivity benefit of collaboration. Our results also suggest that a group's physiological composition, like its collective intelligence, is important to consider.

6.1 Implications

Our results have important implications for productivity, intelligent scheduling, experimental design, and diversity.

Recent research on microproductivity [54] has focused on ways to leverage the efficiency of microtasks to reduce the burden of ramp up and wind down around working hours [65] and to intelligently coordinate group writing [55]. Current conceptualizations of individual and group productivity have overlooked the importance of daily physiological rhythms that cannot be consciously controlled. Given prior research on the impact of circadian timing on individual cognition and now the present work on group cognition and creativity, there is growing evidence that physiology is a pivotal driver in productivity. Microproductivity systems for collaboration could therefore personalize task allocation based on physiology. Straightforward microtasks, such as editing grammatical and spelling mistakes, could be scheduled at users' peak circadian arousal times. On the other hand, imaginative tasks, such as sketching a prototype, could be scheduled at users' off-peak times. As such, circadian rhythm-based models of productivity could augment existing context-aware approaches in microproductivity.

Popular group scheduling tools such as Doodle [23] and WhenIsGood [62] rely on majority voting to reach a consensus on meeting times. Our results, however, show that such heuristics fall short of empowering groups to fully realize their creative and analytic potential. Groups are more creative during the majority's circadian off-peak time of day but are more analytical during the majority's peak time. We recommend scheduling tools identify group members' chronotypes at the point of entering availabilities (without sensors but with a few survey questions), consider the meeting goals (analytic vs. creative), and then rank order times based on the majority's chronotype and availability to schedule group collaboration at the optimal time. More broadly, we argue that physiologically informed, novel organizational structures and tools could more effectively support group creativity.

Research with small groups in HCI and CSCW have been conducted when studies are opportune (e.g., during already scheduled meetings) or scheduled for logistical ease. However, group cognition changes over the course of the day based on group members' chronotypes and circadian arousal. Therefore, observations and the impact of any intelligent systems may be modulated by group members' circadian rhythms and the time of testing. Our findings show that timing could be everything. Researchers should therefore consider the kinds of cognitive load and tasks they expect group members to complete, and if the tasks are analytic or creative, schedule studies during the most appropriate circadian times (and likely both peak and off-peak hours).

Although organizational diversity efforts have primarily focused on age, gender, and ethnicity, we find that temporal diversity is a new factor to consider. Working with people who are in the office at the same time may be convenient, but doing so will likely create a group that is homogeneous in chronotype. Based on our findings, such a group is likely to experience greater physiological synchrony and social cohesion but underperform compared to groups with diverse chronotypes. The exact mechanism by which circadian diversity improves group work is unknown, but the benefits of temporal diversity are noticeable throughout the day. A physiological, temporal dimension to group diversity is an important perspective to consider in group work.

7 LIMITATIONS AND FUTURE WORK

The present work brings attention to the importance of physiology in small group work and opens up opportunities for the invention of novel, physiologically-aware and driven systems to support group formation and work. To realize this possibility, limitations in the present work should be addressed with future research.

The participants in our study were employees of a technology company, so there is a concern that our sample was biased towards people who may have increased aptitude towards the incremental and insight problems we used in our experiment. However, our participants had diverse occupations, including software engineering, project management, and research, and the impact of circadian rhythms has been shown to be basic and pervasive, independent of skill and intelligence. For increased external validity, a replication of this experiment should be conducted with a different population. Future reproductions of this study should also control for participants' sleep patterns. Given the early start time of the morning session, it is possible that morning chronotypes had higher sleep quality and/or slept longer than evening chronotypes who may have stayed up late the night before. For increased coverage of participants' circadian rhythms, future studies should also consider adding a study session towards the middle of the day.

The experimental tasks we used in our study were based on psychological taxonomies of problem solving [35]. Based on prior work [20, 22], we operationalized convergent creativity to be the ability to derive novel and useful solutions to problems that required cognitive re-representation. Although individual subjective ratings have been used to validate and distinguish insight problems from other kinds of problems [60], we did not gather ratings from groups because we wanted to focus groups' attention and time on the tasks rather than ask for individual ratings or require groups to collectively agree on insight ratings after each problem. Future work could incorporate such meta-cognitive rating tasks to compare how individual ratings change throughout the day depending on chronotype and group composition.

The cognitive underpinnings that differentiate insight problems, incremental problems, and other possible problem types are still debated, so reproducing the results here with additional problems related to creativity would provide complementary perspectives on circadian effects on creativity more generally. We chose to study insight problems as a first step towards understanding physiological effects of creativity. To study how circadian rhythms impact creativity in context, future field experiments are necessary. For instance, researchers could ask homogeneous and heterogeneous groups to schedule meetings at peak and off-peak hours and then observe project outcomes in the longer term. We hope to integrate our findings into systems for intelligent scheduling to further test the impact of circadian cycles on group work.

In this paper, we did not take genetic makeup into account even though genotypes may affect chronotypes [48] because we did have access to participants' genotype information. We conducted the study during two weeks in the late summer that did not span the solstice during which daylight duration changes dramatically, so seasonal effects on chronotypes are considered random effects in our models [48].

We measured participants' electrodermal activity and heart rate using wrist-worn sensors. We would have liked to capture heart rate variability (HRV) as well, but the signals collected using the wrist-worn sensors were too noisy for calculating HRV. Even though participants wore the sensors on their non-dominant hands, they moved their hands frequently during problem solving tasks. Future studies could consider measuring HRV using sensors that are slightly more invasive than wrist-worn sensors, such as chest straps [46].

Future work should also examine what interventions could be made to either exaggerate time of day differences on creativity or reduce these differences so that groups can be optimally creative even during their off-peak hours. There may be ways to leverage team members who are in their off-peak times to take on more leadership roles during creative meetings scheduled during these times. Other possible interventions include lighting changes to meeting rooms or interfaces used for computer-mediated systems. For instance, researchers have previously found that the presence of a blue light increased individuals' convergent creative thinking [1]. Future research could examine the impact of blue light in rooms, interfaces, and group members' chronotypes on group creativity.

8 ACKNOWLEDGEMENTS

We are grateful to our anonymous reviewers; the participants in our study who volunteered their time and energy; Scott Counts and Gloria Mark for their input on the statistical reporting of our results; Bongshin Lee, Dan Marshall, Arpita Bhattacharya, Everlyne Kimani, and Deepali Aneja for help conducting the in-lab studies; Austin Hutchin, Duane Sawyer, Stephen Mickey and their team who helped recruit participants for the study; Shuk Kam and Gregory Lee for their support with hardware; the MSR Ethics team for their support of the study; and Kaitlin Smith and Douglas Granger at the University of California, Irvine's Institute for Interdisciplinary Salivary Bioscience Research for their insight and guidance about salivary physiology measures that could impact chronotype and cognition.

REFERENCES

[1] Saeed Abdullah, Mary Czerwinski, Gloria Mark, and Paul Johns. 2016. Shining (blue) light on creative ability. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 793–804.

[2] Khan Academy. [n. d.]. Algebra basics: Systems of equations. https://www.khanacademy.org/math/algebra-basics/alg-basics-systems-of-equations, last accessed September 20, 2018.

[3] Adobe. [n. d.]. 99U Magazine. https://99u.adobe.com/.

[4] Modupe Akinola. 2010. Measuring the pulse of an organization: Integrating physiological measures into the organizational scholar's toolbox. *Research in Organizational Behavior* 30 (2010), 203–223.
[5] The New York Times Crossword Archive. [n. d.]. https://www.nytimes.com/crosswords/, last accessed August 9, 2018.
[6] Ivan K Ash and Jennifer Wiley. 2002. Ah-ha, I knew it all along: Differences in hindsight bias between insight and algebra problems. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 24.
[7] Nicola L Barclay and Andriy Myachykov. 2017. Sustained wakefulness and visual attention: moderation by chronotype. *Experimental brain research* 235, 1 (2017), 57–68.
[8] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series.. In *KDD workshop*, Vol. 10. Seattle, WA, 359–370.
[9] Galen V Bodenhausen. 1990. Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science* 1, 5 (1990), 319–322.
[10] Diane B Boivin, Charles A Czeisler, Derk-Jan Dijk, Jeanne F Duffy, Simon Folkard, David S Minors, Peter Totterdell, and James M Waterhouse. 1997. Complex interaction of the sleep-wake cycle and circadian phase modulates mood in healthy subjects. *Archives of general psychiatry* 54, 2 (1997), 145–152.
[11] Wolfram Boucsein. 2012. *Electrodermal activity*. Springer Science & Business Media.
[12] Alan Bränzel, Christian Holz, Daniel Hoffmann, Dominik Schmidt, Marius Knaust, Patrick Lühne, René Meusel, Stephan Richter, and Patrick Baudisch. 2013. GravitySpace: tracking users and their poses in a smart room using a pressure-sensing floor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 725–734.
[13] C Cajochen, K Kräuchi, and A Wirz-Justice. 2003. Role of melatonin in the regulation of human circadian rhythms and sleep. *Journal of neuroendocrinology* 15, 4 (2003), 432–437.
[14] Prerna Chikersal, Maria Tomprou, Young Ji Kim, Anita Williams Woolley, and Laura Dabbish. 2017. Deep Structures of Collaboration: Physiological Correlates of Collective Intelligence and Group Satisfaction.. In *CSCW*. 873–888.
[15] Charles A Czeisler, Michael P Johnson, Jeanne F Duffy, Emery N Brown, Joseph M Ronda, and Richard E Kronauer. 1990. Exposure to bright light and darkness to treat physiologic maladaptation to night work. *New England Journal of Medicine* 322, 18 (1990), 1253–1259.
[16] Amory H Danek, Joshua Williams, and Jennifer Wiley. 2018. Closing the gap: connecting sudden representational change to the subjective Aha! experience in insightful problem solving. *Psychological research* (2018), 1–9.
[17] Janet E Davidson and Robert J Sternberg. 1984. The role of insight in intellectual giftedness. *Gifted child quarterly* 28, 2 (1984), 58–64.
[18] Michael E Dawson, Anne M Schell, and Diane L Filion. 2007. The electrodermal system. *Handbook of psychophysiology* 2 (2007), 200–223.
[19] Edward De Bono. 1967. *The Use of Lateral Thinking*. Penguin Books Harmondsworth,, UK.
[20] Edward De Bono. 1995. Serious creativity. *The Journal for Quality and Participation* 18, 5 (1995), 12.
[21] Julie Delpouve, Rémy Schmitz, and Philippe Peigneux. 2014. Implicit learning is better at subjectively defined non-optimal time of day. *Cortex* 58 (2014), 18–22.
[22] Roger L Dominowski. 1995. Productive Problem Solving Roger L. Dominowski. *The creative cognition approach* 63 (1995).
[23] Doodle. [n. d.]. https://doodle.com/, last accessed September 19, 2018.
[24] Empatica. [n. d.]. https://www.empatica.com/en-eu/research/e4/,lastaccessedSeptember19,2018.
[25] James Fixx. 1972. More Games for the Superintelligent. *New York Public Library* (1972).
[26] Simon Folkard and Timothy H Monk. 1985. Hours of work. (1985).
[27] Simon Folkard and Philip Tucker. 2003. Shift work, safety and productivity. *Occupational medicine* 53, 2 (2003), 95–101.
[28] Lynn Hasher, Rose T Zacks, and Cynthia P May. 1999. Inhibitory control, circadian arousal, and age. (1999).
[29] Jim A Horne and Olov Östberg. 1976. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology* (1976).
[30] Michael J Hove and Jane L Risen. 2009. It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition* 27, 6 (2009), 949–960.
[31] Regan L Mandryk and Kori M Inkpen. 2004. Physiological indicators for the evaluation of co-located collaborative play. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 102–111.
[32] Cynthia P May. 1999. Synchrony effects in cognition: The costs and a benefit. *Psychonomic Bulletin & Review* 6, 1 (1999), 142–147.
[33] Cynthia P May and Lynn Hasher. 1998. Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance* 24, 2 (1998), 363.
[34] Cynthia P May, Lynn Hasher, and Ellen R Stoltzfus. 1993. Optimal time of day and the magnitude of age differences in memory. *Psychological Science* 4, 5 (1993), 326–330.

[35] Janet Metcalfe and David Wiebe. 1987. Intuition in insight and noninsight problem solving. *Memory & cognition* 15, 3 (1987), 238–246.
[36] Dan Mønster, Dorthe Døjbak Håkonsson, Jacob Kjær Eskildsen, and Sebastian Wallot. 2016. Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology & behavior* 156 (2016), 24–34.
[37] Martin C Moore-Ede. 1986. Physiology of the circadian timing system: predictive versus reactive homeostasis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 250, 5 (1986), R737–R752.
[38] Greg Murray, Nicholas B Allen, and John Trinder. 2002. Mood and the circadian system: Investigation of a circadian component in positive affect. *Chronobiology international* 19, 6 (2002), 1151–1169.
[39] Greg Murray and Allison Harvey. 2010. Circadian rhythms and sleep in bipolar disorder. *Bipolar disorders* 12, 5 (2010), 459–472.
[40] Greg Murray, Christian L Nicholas, Jan Kleiman, Robyn Dwyer, Melinda J Carrington, Nicholas B Allen, and John Trinder. 2009. NatureâĂŹs clocks and human mood: The circadian system modulates reward motivation. *Emotion* 9, 5 (2009), 705.
[41] Reiner Nikula. 1991. Psychological correlates of nonspecific skin conductance responses. *Psychophysiology* 28, 1 (1991), 86–90.
[42] Kati Nowack and Elke Van Der Meer. 2018. The synchrony effect revisited: chronotype, time of day and cognitive performance in a semantic analogy task. *Chronobiology international* 35, 12 (2018), 1647–1662.
[43] Daniel Olguín Olguín, Benjamin N Waber, BN Taemie Kim, Akshay Mohan, Koji Ara, and Alex Pentland. 2008. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. Institute of Electrical and Electronics Engineers.
[44] Suzanne J Peterson, Christopher S Reina, David A Waldman, and William J Becker. 2015. Using physiological methods to study emotions in organizations. In *New ways of studying emotions in organizations*. Emerald Group Publishing Limited, 1–27.
[45] Thomas V Petros, Bill E Beckwith, and Maureen Anderson. 1990. Individual differences in the effects of time of day and passage difficulty on prose memory in adults. *British Journal of Psychology* 81, 1 (1990), 63–72.
[46] Daniel J Plews, Ben Scott, Marco Altini, Matt Wood, Andrew E Kilding, and Paul B Laursen. 2017. Comparison of heart-rate-variability recording with smartphone photoplethysmography, Polar H7 chest strap, and electrocardiography. *International Journal of Sports Physiology and Performance* 12, 10 (2017), 1324–1328.
[47] Frand Restle and James H Davis. 1962. Success and speed of problem solving by individuals and groups. *Psychological Review* 69, 6 (1962), 520.
[48] Kathryn A Roecklein, Patricia M Wong, Peter L Franzen, Brant P Hasler, W Michael Wood-Vasey, Vishwajit L Nimgaonkar, Megan A Miller, Kyle M Kepreos, Robert E Ferrell, and Stephen B Manuck. 2012. Melanopsin gene variations interact with season to predict sleep onset and chronotype. *Chronobiology international* 29, 8 (2012), 1036–1047.
[49] Jonathan W Schooler and Joseph Melcher. 1995. The ineffability of insight. (1995).
[50] Jonathan W Schooler, Stellan Ohlsson, and Kevin Brooks. 1993. Thoughts beyond words: When language overshadows insight. *Journal of experimental psychology: General* 122, 2 (1993), 166.
[51] Michael H Smolensky and Gilbert E D'alonzo. 1993. Medical chronobiology: concepts and applications. *American Review of Respiratory Disease* 147 (1993), S2–S2.
[52] Robert J Sternberg and Janet E Davidson. 1982. The mind of the puzzler. *Psychology Today* 16, 6 (1982), 37–44.
[53] Chiew Seng Sean Tan, Johannes Schöning, Kris Luyten, and Karin Coninx. 2014. Investigating the effects of using biofeedback as visual stress indicator during video-mediated collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 71–80.
[54] Jaime Teevan. 2016. The future of microwork. *XRDS: Crossroads, The ACM Magazine for Students* 23, 2 (2016), 26–29.
[55] Jaime Teevan, Shamsi T Iqbal, and Curtis Von Veh. 2016. Supporting collaborative writing with microtasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2657–2668.
[56] Kenneth J Travers, L C Dalton, V F Bruner, and A R Taylor. 1976. *Using advanced algebra*. Doubleday Toronto.
[57] Tanya Vacharkulksemsuk and Barbara L Fredrickson. 2012. Strangers in sync: Achieving embodied rapport through shared movements. *Journal of experimental social psychology* 48, 1 (2012), 399–402.
[58] Rick B Van Baaren, Rob W Holland, Kerry Kawakami, and Ad Van Knippenberg. 2004. Mimicry and prosocial behavior. *Psychological science* 15, 1 (2004), 71–74.
[59] Ruth Wageman, J Richard Hackman, and Erin Lehman. 2005. Team diagnostic survey: Development of an instrument. *The Journal of Applied Behavioral Science* 41, 4 (2005), 373–398.
[60] Margaret E Webb, Daniel R Little, and Simon J Cropper. 2016. Insight is not in the problem: Investigating insight in problem solving across task types. *Frontiers in psychology* 7 (2016), 1424.
[61] Margaret E Webb, Daniel R Little, and Simon J Cropper. 2018. Once more with feeling: Normative data for the aha experience in insight and noninsight problems. *Behavior research methods* 50, 5 (2018), 2035–2056.

[62] WhenIsGood. [n. d.]. http://whenisgood.net/, last accessed September 19, 2018.
[63] Mareike Wieth and Bruce D Burns. 2006. Incentives improve performance on both incremental and insight problem solving. *The Quarterly Journal of Experimental Psychology* 59, 8 (2006), 1378–1394.
[64] Mareike B Wieth and Rose T Zacks. 2011. Time of day effects on problem solving: When the non-optimal is optimal. *Thinking & Reasoning* 17, 4 (2011), 387–401.
[65] Alex C Williams, Harmanpreet Kaur, Gloria Mark, Anne Loomis Thompson, Shamsi T Iqbal, and Jaime Teevan. 2018. Supporting Workplace Detachment and Reattachment with Conversational Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 88.
[66] Scott S Wiltermuth and Chip Heath. 2009. Synchrony and cooperation. *Psychological science* 20, 1 (2009), 1–5.
[67] Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science* 330, 6004 (2010), 686–688.
[68] Carolyn Yoon. 1997. Age differences in consumers' processing strategies: An investigation of moderating influences. *Journal of Consumer Research* 24, 3 (1997), 329–342.

9 APPENDIX A: PROBLEM SOLVING MATERIALS

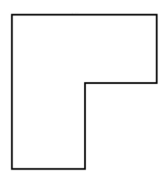
9.1 Insight problems

Chain: A woman has 4 pieces of chain. Each piece is made up of 3 links. She wants to join the pieces into a single closed loop of chain. To open a link costs 2 cents and to close a link costs 3 cents. She only has 15 cents. How does she do it? (originally from [19])

Prisoner: A prisoner was attempting to escape from a tower. He found in his cell a rope which was half long enough to permit him to reach the ground safely. He divided the rope in half and tied the two parts together and escaped. How could he have done this? Explain specifically what he did. (originally from [47])

Water lilies: Water lilies double in area every 24 hours. At the beginning of summer there is one water lily on the lake. It takes 60 days for the lake to become completely covered with water lilies. On which day is the lake half covered? (originally from [52])

Divide figure: Show how you can divide this figure into 4 equal parts that are the same size and shape. (originally from [25])



Coin: A dealer in antique coins got an offer to buy a beautiful bronze coin. The coin had an emperor's head on one side and the date 544 BC stamped on the other. The dealer examined the coin, but instead of buying it, he called the police. Why? (originally from [64])

Socks: If you have black socks and brown socks in your drawer, mixed in a ratio of 4 to 5, how many socks will you have to take out to make sure that you have a pair the same color? (originally from [52])

Pennies: Show how you can arrange 10 pennies so that you have 5 rows (lines) of 4 pennies in each row. (originally from [25])

Magician: A magician claimed to be able to throw a ping pong ball so that is would go a short distance, come to a dead stop, and then reverse itself. He also added that he would not bounce the ball against any object or tie anything to it. How could he perform this feat? (originally from [3])

Weights: There are ten bags, each containing ten weights, all of which look identical. In nine of the bags each weight is 16 ounces, but in one of the bags the weights are actually 17 ounces each. How is it possible, in a single weighing on an accurate weighing scale, to determine which bag contains the 17-ounce weights? (originally from [3])

Tree planting: A landscape gardener is given instructions to plant 4 special trees so that each one is exactly the same distance from each of the others. How is he able to do it? (originally from [19])

Captain: Captain Scott was out for a walk when it started to rain. He did not have an umbrella and he wasn't wearing a hat. His clothes were soaked yet not a hair on his head got wet. How could this happen? (originally from [3])

Horse: A man bought a horse for $60 and sold it for $70. Then bought it back for $80 and sold it for $90. How much money did he make in the horse business? (originally from [19])

9.2 Incremental problems

Card: Three cards from an ordinary deck are lying on a table, face down. The following information is known about those three cards (all the information refers to the same three cards):

- To the left of a Queen, there is a Jack.
- To the left of a Spade, there is a Diamond.
- To the right of a Heart, there is a King.
- To the right of a King, there is a Spade.

Can you assign the proper suit to each picture card? (originally from [63])

Crossword puzzle: Adapted from The New York Times Daily Crossword archive. Originally published on Wednesday, July 4, 2017. Access was free of charge on August 9, 2018 [5]

Age: Ann is twice as old as her son. They were both born in June. Ten years ago Ann was three times as old as her son. What are their present ages? (originally from [63])

Store: Smith is a butcher and president of the street storekeepers' committee, which also includes the grocer, the baker, and the pharmacist. They all sit around a table.

- Smith sits on Jones' left.
- Davis sits at the grocer's right.
- Bailey, who faces Jones, is not the baker.

Assign each storekeeper to the correct store. (originally from [63])

Fill in the blank: 2, 4, 6, 30, 32, 34, 36, 40, 42, 44, 46, 50, 52, 54, 56, 60, 62, 64, 66, ___? (originally from [3])

Algebra: Solve for x, y, and z:

x + y - 3z = -10
x - y +2z = 3
2x + y - z = -6

(inspired by question in [35], adapted from [2])

Bachelor: Five bachelors, Andy, Bill, Carl, Dave, and Eric, go out together to eat five evening meals (Fish, Pizza, Steak, Tacos, and Thai) on Monday through Friday. It was understood that Eric would miss Friday's meal due to an out of town wedding. Each bachelor served as the host at a restaurant of his choice on a different night. The following information is known:

- Carl hosted the group on Wednesday.
- The fellows ate at a Thai restaurant on Friday.
- Bill, who detests fish, volunteered to be the first host.
- Dave selected a steak house for the night before one of the fellows hosted everyone at a raucous pizza parlor.

Which bachelor hosted the group each night and what food did he select? (originally from [63])

Puzzle: If the puzzle you solved before you solved this one was harder than the puzzle you solved after you solved the puzzle you solved before you solved this one, was the puzzle you solved before you solved this one harder than this one? (originally from [47])

Solve for m:

$$\frac{m-3}{2m} - \frac{m-2}{2m+1} = 0$$

(originally from [56])

Flowers: Four women, Anna, Emily, Isabel, and Yvonne, receive a bunch of flowers from their partners, Tom, Ron, Ken, and Charlie. The following information is known:

- Anna's partner, Charlie, gave her a huge bouquet of her favorite blooms; which aren't roses.
- Tom gave daffodils to his partner (not Emily).

- Yvonne received a dozen lilies, but not from Ron.

What type of flowers (carnations, daffodils, lilies, or roses) were given to each woman and who is her partner? (originally from [63])

Words: Rearrange the following patterns to make familiar words:

- runghy
- flymia
- mulcica
- dornev
- lendraca

(originally from [3])

Card game: Three people play a game in which one person loses and two people win each round. The one who loses must double the amount of money that each of the other two players has at that time. The three players agree to play three games. At the end of the three games, each player has lost one game and each person has $8. What was the original stake of each player? (originally from [35])

Received April 2019; revised June 2019; accepted August 2019