Psychophysiological Baseline Methods and Usage

Avonie Parchment, Ryan W. Wohleber(✉), and Lauren Reinerman-Jones

Institute for Simulation and Training, University of Central Florida, Orlando, FL, USA
{aparchme,rwohlebe,lreinerm}@ist.ucf.edu

Abstract. There are several different baseline techniques available for completing psychophysiological research, yet no overarching set of guidelines exists to help researchers choose the best method. This review examines several methods used in various fields and highlights the importance and pitfalls of each. As part of this effort we conducted a small study that examines three different baseline techniques. In line with the Law of Initial Value (LIV), outcomes signal a strong positive effect for measures when utilizing a resting baseline, a weaker positive effect when utilizing a baseline directly before tasking, and a nominal effect when calibrating using a comprehensive baseline. The authors caution future researchers to fully assess the needs of their experiment before utilizing comprehensive, vanilla, or resting baselines, and to weigh the consequences of the length and number of baselines utilized. Further investigation of low workload and vigilance tasking is needed to determine whether use of vanilla and comprehensive baselines provide better contrast than a resting baseline.

Keywords: Resting baseline · Comprehensive baseline · Vanilla baseline · Methods · Psychophysiological measures

1 Introduction

In Human-Computer Interaction, Human Factors Psychology, Neurophysiology, and related fields, psychophysiological measures have become a pillar for research into the human state. For example, recent efforts have employed psychophysiological sensors to create human-robot closed loop systems, [1] test the effectiveness of games [2], and to understand how workload and stress influence performance under various conditions [3]. Psychophysiological measures provide a direct gauge of human state, in contrast to subjective measures, which rely on introspection, and performance measures, which must infer state from behavioral outcomes. As these measures have become increasingly more accessible to researchers due to improvements in technology and reduction in cost, best practices for utilizing psychophysiological measures need to be evaluated. Notably, the use of baselines is an important part of these methods.

Psychophysiology has well over a century of history [4], but is a fairly new formal discipline. A foundational principle of psychophysiology is Wilder's Law of Initial Value (LIV) [5] which stipulates that the direction of a psychophysiological response depends on initial state. If a person's initial arousal is high, then a task designed to elevate arousal may only show a modest increase, if any, when contrasted with the initial state. However, if a person's initial arousal is low, the effects of the same task may appear

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC 2016, Part I, LNAI 9743, pp. 361–371, 2016.
DOI: 10.1007/978-3-319-39955-3_34

larger and discernable when contrasted to the initial state [4]. This phenomenon is seen in investigations concerning heart rate, respiration rate, and skin resistance (SR), but not in skin conductance (SC) or temperature [4]. In order to best understand this phenomenon and how it effects psychophysiological assessment, an evaluation of various baseline methodologies should be undertaken.

Over the decades, differences in measurement (c.f., [6–9]) have arisen, yet despite the large volume of psychophysiological research conducted over the years, little work has been done to systematically evaluate the available baseline options. From the reports of the procedures used by various researchers, it is evident that these procedures differ depending on the researcher's paradigm, field of study, and experimental content. Unfortunately, these differences in baseline practices have confounded replication efforts as well as efforts to improve methodological approaches [7, 9]. At the root of these issues is the lack of established common practice rooted in sound empirical investigation [4]. Therefore, an examination of baselines for psychophysiological measures is sought and such questions to answer include:

1. What kind of baseline is needed?
2. How long should a baseline be?
3. Where should the baseline be placed?

Answering these questions is the first step to realizing the full potential of psychophysiological measures. The following sections provide a basis for answering these questions and include results from a small investigation which was undertaken to illustrate the impact of baseline choice on the interpretation of psychophysiological outcomes.

2 Types of Baselines

There are several types of baselines for psychophysiological measurement. Among the most prominent are the basal/resting, vanilla, and comprehensive baselines. Within each type, there are multiple variations which can contribute to the difficulties with replicability, validity, and even the ability to generate valid conclusions. These baseline variations are described in subsequent sections and our discussion will include the benefits and shortcomings of each.

2.1 Basal/Resting Baselines

Basal or tonic activity refers to the resting level of activity for a psychophysiological measure. A true basal baseline refers to the absolute lowest reactive state of a participant [10]. Generally, the procedure includes a reduced or simple diet to control any changes in psychophysiological responses due to consumed items. A participant arrives at the testing location and, after being connected to one sensor, is monitored in a supine position that could extend several hours [11]. The benefit of this type of baseline is the use of a highly controlled environment where any variability is due entirely to the participant. However, this baseline also runs the risk of boring the participant, putting the participant to sleep, or exciting the participant by putting him or her in an area devoid of human interaction [8].

While highly controlled, basal baselines enhance and emphasize the individualized responses between participants, making general conclusions based on a sample difficult. Further, these baselines are impractical for obtaining a pre-task resting state.

The shorter resting baseline circumvents the problem of increased individual variability between subjects produced by the basal baseline. By reducing the length of the baseline to ten minutes or less [7], keeping the participant upright but sitting, and requesting that the participant keep his or her eyes open [12], boredom and drowsiness is decreased and the variability between participants is reduced. Although this resting type of baseline addresses some of the problems found in basal baseline procedures, participants still have ample opportunity to lose concentration and let their minds wander; instructions for resting baselines typically do not stipulate what to think about or how to breathe during the baseline (e.g., a participant might increase her heart rate by either thinking of something unpleasant or by taking quick, short breaths). Though a few researchers have been able to determine at which point a baseline stabilizes [7], there remains sizable variability even within the duration of resting baselines. The challenge with using resting/basal baselines is the large variability across participants as well as within the participant that is associated with these baselines.

2.2 Vanilla Baselines

In order to reduce variability between participants when comparing baseline measurements, researchers developed the vanilla baseline. For this baseline, participants are given a low task load activity to complete while connected to a psychophysiological sensor. This procedure involves giving each participant the same activity to think about, resulting in reduced variability between participants [7]. Additionally, Piferi et al. [9] found that systolic and diastolic blood pressure when watching a relaxing video was significantly lower than that during a resting recovery period.

While this procedure is considered a way to standardize participants' psychophysiological response, there can still be substantial variability between studies because researchers can select different low task load activity options that may not be comparable to each other. Many researchers are partial to card sorting tasks or other low task load exercises, while others will use a calming video or a number of shapes on a screen [9]. Because of this variability in vanilla baseline task, differences in mean response to the baseline task may not be comparable across experiments. Nonetheless, the vanilla baseline is able to restrict the range of response in a baseline within a study and can induce a relaxed state in participants that is comparable to that of a true resting baseline.

2.3 Comprehensive Baselines

It has been argued that initial baselines alone, of any kind, are inadequate due to their inability to report the true normal psychophysiological state of the participant [13]. Further, researchers have suggested that the response to some stimuli could be overstated if only a resting value is used for comparison [13, 14]. In order to have a more reliable comparison and to gauge whether a meaningful response to a task has truly occurred, it may be advantageous to use the average response to a variety of stimuli, that is, a

comprehensive baseline [13]. Although a comprehensive baseline could refer to a short regimen of various tasks such as that provided by ABM's B-Alert software, for our purposes, we refer specifically to the method of averaging response of all tasks over an entire experimental session, described by Fishel and colleagues as a "gold standard self-calibration period" [13]. By recording and averaging a broader range of responses a participant can have, a researcher may approximate a participant's true average state. Any significant deviation from this comprehensive baseline could be interpreted as a more genuine response to the task in question than would be attained from an arguably artificial resting state.

Despite capturing the full range of responses during an experimental session, comprehensive baselines introduce a large amount of variance in the baseline used for comparison to the experimental task. However, this greater variance may only be seen within the participant response, rather than between participants. Comparisons between the different types of baselines could determine if comprehensive baselines may reduce variability between participants.

The comprehensive baseline highlights important questions for psychophysiological research: is the comparison between a response to a task and an initial resting baseline artificial because participants' psychophysiological response is artificially attenuated to achieve contrast? Typical resting baseline procedures use behavior that may not be a normal part of everyday experience. Additionally, if such procedures are artificial, is the comparative response to subsequent tasking a valid indicator of response to a task? Although there may not be any definitive answer to these questions, a comparison of resting, vanilla, and comprehensive baselines may provide some insights.

3 Baseline Length and Number of Baselines

In addition to the type of baseline, the importance of the length and number of baselines recorded during an experimental session should also be noted. As mentioned in our discussion on resting baselines, a researcher chooses the baseline length based on how long the researcher expects it to take for participants' psychophysiology response to stabilize. The initial portion of a baseline will inevitably contain greater variability than subsequent portions as the participant acclimates to the baseline task. After some time, the participant's psychophysiology reaches some stable state. The length of time required for this acclimation process, and thus the baseline, is often unclear due to the number of different psychophysiological measures, the type of experiment, and the type of equipment.

The use of baselines in-between tasks has also been a matter of debate in the research community. Gauging of psychophysiological response is reliant on contrast with some initial state (c.f., LIV). Adding resting periods between tasking can allow the participant's psychophysiology to return to this initial level and enable researchers to obtain unadulterated assessments of participant's response to multiple tasks [15]. Unfortunately, the introduction of additional baselines may also elicit restlessness, boredom, and mind wandering that results in the corruption of the psychophysiological measurement.

3.1 Length of Baseline

Baseline research is marked by a lack of consensus regarding the length of time for a baseline to stabilize [6, 11, 12, 15, 16]. In the studies just listed, resting baseline length for a heart rate measure ranged from 8 to 15 min; the method for capturing participants' heart rate differed depending on experimental goals and context. Of several methods, Fishel and colleagues [13] used a moving 2 min window when calculating baseline. Jennings and colleagues [7] had a more sensitive apparatus and let their participants rest for over 25 min at the start of the experimental session. They then took the entire time of this baseline and graphed it, noting the time where variability was minimized. While all these methods may be valid, it is important to recognize that instead of standard practices that all researchers must adhere to, there exists a plethora of available baseline practices and generous flexibility of each procedure's parameters (e.g., length), which allows researchers to choose methods that offer them the best chance of showing an effect of experimental manipulations. It may be prudent to consider whether or not such flexibility in practice is advantageous to inquiry or complicates the search for reliable effects.

3.2 Number of Baselines

As mentioned above, some researchers choose to put resting periods (i.e., baselines) in-between tasking in order to reset psychophysiological responses before exposure to new tasking or experimental manipulations. Jennings and colleagues [7] suggest that only a few minutes is required to shed the influence of previous tasking. It might be argued that questionnaires between tasks are sufficient to bring participants back to baseline levels. Others might argue that baselines taken immediately prior to a task provides the most appropriate initial value from which to gauge response to a stimulus [15]. The different perspectives on multiple resting period practice can be summarized thusly: Baselines between tasking are needed to bring psychophysiological response back to the initial baseline levels. Baselines directly before tasking are necessary to gauge true response to a task. Finally, rest times between tasking is irrelevant to psychophysiological response as all psychophysiological responses are relative to some internal constant, basal level. While several articles touch upon these positions [6, 8, 13], no direct comparison of each of these possibilities yet exists.

3.3 Baseline Investigation Summary

Investigation of baseline methods is necessary for improving understanding of psychophysiological response and promoting more generalizable practices that allow for comparison both within and across programs of research. In our review of past baseline research, we identified possible ramifications of varying baseline methods, baseline lengths, and number of baselines. The following sections detail a preliminary effort to help alleviate concerns with baseline practices.

4 Experimental Approach

The present research investigates the differences between different types of baselines: resting and comprehensive. This effort is intended to generate additional inquiry into baseline research. The present experiment follows some of the procedures used by Fishel and colleagues [13] who also investigated multiple baselines types. Specifically, we assessed an initial resting baseline which came first in the experimental procedure and a comprehensive baseline was calculated using every data point throughout the experiment (Fishel and colleague's "gold standard"). Due to the copious number of vanilla baseline methods (calming video, card sorting task, listening to instructions), the authors felt that including a vanilla baseline within this investigation would be beyond the digestible scope of the present research and will be the focus of future research. The goal of this investigation was to compare the two baselines representing the low arousal (initial resting and resting directly before task baselines) and high arousal (comprehensive or gold standard baselines) extremes in a range of baseline options in order to understand the implications of each for different types of research questions and task manipulations [13, 16].

Hypothesis 1: Psychophysiological Responses Across the Various Tasks Would be Different. The materials chosen for this investigation represent a range of possible tasking in human performance research. One task required participants to think about the self and how they would react in different situations. Another required participants to react to changes in a short situation. A third task required them to reflect quickly on intuitive answers. The final task required participants to repeatedly think through a hypothetical scenario and react accordingly to how they expected the scenario to go. Each of these tasks required different processing and were expected to elicit very different responses.

Hypothesis 2: Different Baseline Methods Would Result in Different Initial Value from Which to Compare Psychophysiological Response to Subsequent Manipulations. This investigation compares resting and comprehensive baselines (which tend to indicate different responses to the same task) to the use of a series of baselines taken immediately prior to a task. Fishel and colleagues [13] found that resting baselines showed the greatest bias toward positive response and that a practice (perhaps vanilla) baseline was biased toward negative response. The comprehensive baseline was shown to be a less biased measurement of participant response since it also registered a change in arousal, but did not inflate or deflate responses in relation to other measurement methods. However, a baseline taken directly before tasking was never investigated in relation to these other calibration methods. It was hypothesized that while there would be differences between baseline calibration methods, a baseline taken before the task in question would result in moderate participant response while a resting baseline would show higher response to positive arousal and a comprehensive baseline would show a muted response.

5 Method

5.1 Participants

Seventy-six volunteers from the Central Florida area participated. Due to technical problems with the psychophysiological sensors found after the end of the study, data from two participants were omitted from analysis. Analyses were performed using data from the remaining 74 participants (34 women, 40 men, M_{age}: 21.72 years).

5.2 Materials

Unless otherwise noted, all tasking was administered on a desktop computer.

Everyday Moral Decision Making Task. This task was a mix of two morality measures [17] which asked participants to choose altruistic or egoistic responses to everyday decisions ranging in emotional impact from low to high. Other questions included utilitarian questions that asked participants if they were willing to sacrifice one for the good of many.

Change Detection. A two minute version of the mixed initiative (MIX) testbed [18] asked participants to classify icon changes on a computer screen. The amount of changes varied from the first minute (one change every 8–12 s) to the second (one change every 4 s).

Cognitive Reflection. Fredrick [19] developed the Cognitive Reflection Test that assessed a participant's ability to choose the non-intuitive response to a set of numerical questions.

Paper Game. This task replicates the MIT Beer Game in a computer setting [20] and assesses a participant's ability to make decisions in a supply chain context. The participant played the role of a retailer and had to find a balance between inventory size, orders, and revenue for the entire supply chain over the course of a simulated year.

Electrocardiography (ECG). ECG was monitored using the Advanced Brain Monitoring B-Alert X10 System sampling at 256 Hz. Raw values were Winsorized before analysis. ECG yielded measures of inter-beat interval and heart rate variability, which were recorded using single-lead electrodes placed on the center of the right clavicle and on the lowest left rib.

5.3 Procedure

Participants read an informed consent then completed the initial five minute baseline while keeping their eyes on a dark computer screen and remaining quiet, but alert. Participants then completed pre-questionnaires, the moral decision making task, a resting period, the change detection task, a second resting period, the cognitive reflection task, post task questionnaires, a final resting period, the paper game, and final questionnaires.

6 Results

Each task's percentage difference from baseline was calculated using each of three different baselines: an initial resting baseline taken at the beginning of the study, the baseline immediately prior to the task, and the comprehensive baseline which was the average of all data points in the experiment. The first task's calculation for baseline immediately prior to task used the initial resting baseline.

In order to determine if tasks differed from one another and if baseline methods produced different results from one another (Hypotheses 1 and 2), a 4 (task) by 3 (baseline method) repeated measures ANOVA was run. For psychophysiological response to task, Table 1 shows that for inter-beat interval and for heart rate variability, there was a main effect for task type and baseline calibration method. Planned pairwise comparisons showed that all tasks, with the exception of the comparison between the CRT and the Paper Game, were significantly different from one another for heart rate variability. For heart rate inter-beat interval, all tasks were significantly different from each other with the exception of the change detection task and the Paper Game.

Table 1. Within-subjects 4 (task) × 3 (type of baseline) ANOVA for heartbeat measures

	df	*F*	η_p^2	*p*
HRV				
Baseline type	1.629, 118.906	45.289	.383	<.001
Task	3, 219	30.717	.296	<.001
Type * Task	3.130, 228.463	7.705	.095	<.001
IBI mean				
Baseline type	1.296, 94.618	38.327	.344	<.001
Task	2.425, 177.041	27.832	.276	<.001
Type * Task	2.565, 189.461	23.134	.241	<.001

To interpret the difference in response based on calibration method, each response was graphed. Figure 1 shows the difference in percent change from baseline for heart rate variability. With the exception of the first task, which used the initial baseline as the immediately prior baseline (moral decision making), the baseline taken immediately prior to each task resulted in an apparent response that was weaker than the resting baseline based response but stronger than the comprehensive baseline base response. Planned comparisons showed that each difference was significant at the $p < .001$ level.

Figure 2 shows the difference in percent change from baseline for inter-beat interval. With the exception of the first task which used the initial baseline as the immediately-prior baseline (moral decision making), the baseline taken directly before each task resulted in a response between the resting and comprehensive baselines for the change detection task. However, this difference was not seen for Cognitive Reflection or for the Paper Game. In fact, for these two tasks, the use of a baseline directly before the task resulted in a negative response. Planned comparisons showed that the resting baseline calibration method was significantly different from all others at the $p < .001$ level. For inter-beat interval,

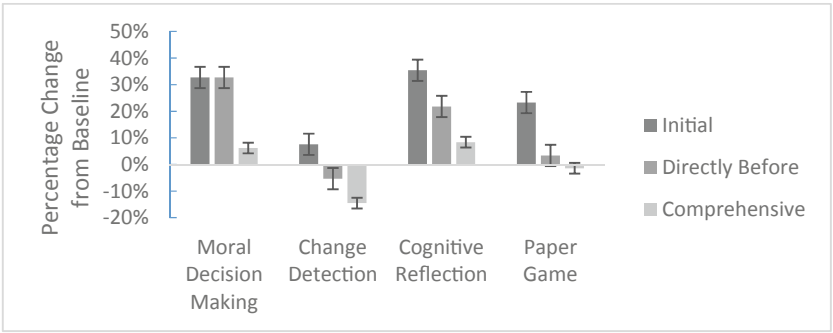

Fig. 1. Heart rate variability percentage change from baseline for each of the four tasks calibrated with the three different methods (Color figure online).

however, the calibration method of taking a baseline directly before the task and the method of taking a comprehensive baseline were not significantly different.

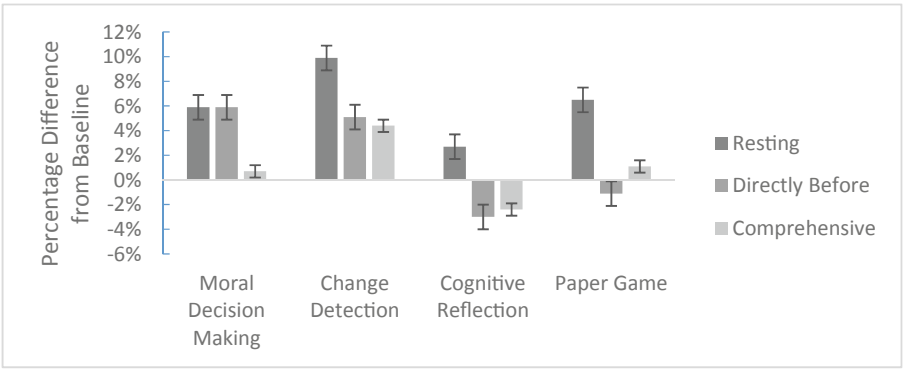

Fig. 2. Inter-beat interval percentage change from baseline for each of the four tasks calibrated with the three different methods (Color figure online).

7 Discussion

The present effort built on that of Fishel and colleagues [13] to determine how baselines taken immediately prior to a task compare to resting and comprehensive baselines. To achieve this goal, each baseline was used to calculate percent change from baseline for four different tasks. The tasks were significantly different from each other with the exception of the Paper Game and the Cognitive Reflection task for heart rate variability and the Paper Game and the change detection task for inter-beat interval. It is possible that the Paper Game shares some aspects of processing with the CRT and the Change Detection Task, though the fact that the rest of the tasks differ was acceptable for the purposes of this research. With these exceptions, we feel that the tasks chosen are

sufficiently different to show a range of tasking which provide some measure of generalizability for the baseline related findings.

We attempted to see if a baseline taken directly before a task resulted in a moderate response to the task in relation to a resting baseline and a comprehensive baseline. For heart rate variability, this moderate response was certainly the observed phenomenon. As hypothesized, for each task, the resting baseline did show a tendency to indicate a positive percent change from baseline for each task, just as the comprehensive baseline would show a muted or negative percent change from baseline. It seems possible that this difference occurred due to a number of factors: the resting baseline is supposedly a measure of rest while the comprehensive baseline is one of average state. Taking a baseline directly before a task seemed to elicit a response exactly in-between the resting and comprehensive baseline calibration responses as hypothesized. However, inter-beat interval calibrated using the baseline taken directly before the task, versus the other two methods, was markedly shorter for the last two tasks (CRT and Paper Game). This difference raises the question of whether previous tasking affected participant response to baseline or not. This difference also brings into question methodologies that do not account for this possibility.

Because the calibrated response task showed a very large increase in inter-beat interval when using a resting baseline while other calibration methods showed less of a change from baseline, it seems possible that the resting baseline is, as Fishel and colleagues pointed out, very susceptible to positive response. However, what we did not expect was just how drastically different responses could be depending on the calibration method. Therefore, regardless of the baseline type and length chosen, it seems the intensity of change is most important for consideration when comparing physiological response between tasks.

8 Conclusion

The purpose of the present paper was to review common baseline practices and highlight the challenges involved in choosing an appropriate baseline method when conducting an experiment. The present research indicated that resting baselines are apt to show a large response to a task and that comprehensive baselines show a more muted response, as seen in previous work [13] and as might be predicted by LIV. However, the present research also showed that for a more moderate indication of response, a baseline taken directly before an experimental task may be prudent for showing the effect of the task and not the compounded effect of the entire experiment up to that point in the session. It may be most important to note that the use of any baseline resulted in clear look at participant response to task; a finding that may not have been apparent if no baseline at all had been used. Further research is needed to compare the many different vanilla baseline methods to those reviewed here. Additional research to investigate the effect each of these baselines had on participant response in relation to time on task and session is also needed. We hope that this preliminary effort demonstrated the important consequences of baseline selection and serves as a caution to future psychophysiological work and encouragement for future investigations into baseline methodology.

References

1. Schirner, G., Erdogmus, D., Chowdhury, K., Padir, T.: The future of human-in-the-loop cyber-physical systems. Computer **1**, 36–45 (2013)
2. Mandryk, R.L.: Physiological measures for game evaluation. Game Usability: Advice from the Experts for Advancing the Player Experience, pp. 207–235 (2008)
3. Abich, J., Matthews, G., Reinerman-Jones, L.: Individual differences in UGV operation: a comparison of subjective and psychophysiological predictors. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 59, no. 1, pp. 741–745. SAGE Publications (2015)
4. Cacioppo, J.T., Tassinary, L.G., Berntson, G.: Handbook of Psychophysiology. Cambridge University Press, Cambridge (2007)
5. Wilder, J.: Stimulus and Response: The Law of Initial Value. Wright, Bristol (1967)
6. Cupini, L.M., Matteis, M., Troisi, E., Sabbadini, M., Bernardi, G., Caltagirone, C., Silvestrini, M.: Bilateral simultaneous transcranial doppler monitoring of flow velocity changes during visuospatial and verbal working memory tasks. Brain **119**(4), 1249–1253 (1996)
7. Jennings, J.R., Kamarck, T., Stewart, C., Eddy, M., Johnson, P.: Alternate cardiovascular baseline assessment techniques: vanilla or resting baseline. Psychophysiology **29**(6), 742–750 (1992)
8. Morcom, A.M., Fletcher, P.C.: Does the brain have a baseline? Why we should be resting a rest. Neuroimage **37**(4), 1073–1082 (2007)
9. Piferi, R.L., Kline, K.A., Younger, J., Lawler, K.A.: An alternative approach for achieving cardiovascular baseline: viewing an aquatic video. Int. J. Psychophysiol. **37**, 207–217 (2000)
10. Stern, R.M., Ray, W.J., Quigley, K.S.: Psychophysiological Recording, 2nd edn. Oxford University Press, Oxford (2001)
11. Gerin, W., Pieper, C., Pickering, T.G.: Anticipatory and residual effects of an active coping task on pre- and post-stress baselines. J. Psychosom. Res. **38**, 139–149 (1994)
12. Reinerman-Jones, L.E., Matthews, G., Langheim, L.K., Warm, J.S.: Selection for vigilance assignments: a review and proposed new direction. Theor. Issues Ergon. Sci. **12**(4), 273–296 (2010)
13. Fishel, S.R., Muth, E.R.: Establishing appropriate physiological baseline procedures for real-time physiological measurement. J. Cogn. Eng. Decis. Making **1**(3), 286–308 (2007)
14. Jacob, R.G., Shapiro, A.P.: Is the effect of stress management on blood pressure just regression to the mean? Homeostasis Health Dis. (1994)
15. Piper, S.K., Krueger, A., Koch, S.P., Mahnert, J., Habermehl, C., Stenbrink, J., Obrig, H., Schmitz, C.H.: A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. Neuroimage **85**(1), 64–71 (2014)
16. Stroobant, N., Vingerhoets, G.: Transcranial Doppler ultrasonography monitoring of cerebral hemodynamics during performance of cognitive tasks: a review. Neuropsychol. Rev. **10**(4), 213–231 (2000)
17. Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D.: The neural bases of cognitive conflict and control in moral judgment. Neuron **44**(2), 389–400 (2004)
18. Barber, D., Leontyev, S., Sun, B., Davis, L., Nicholson, D., Chen, J.Y.: The mixed-initiative experimental testbed for collaborative human robot interactions. In: Collaborative Technologies and Systems, IEEE, pp. 483–489 (2008)
19. Frederick, S.: Cognitive reflection and decision making. J. Econ. Perspect. **19**, 25–42 (2005)
20. Kaminsky, P., Simchi-Levi, D.: A new computerized beer game: a tool for teaching the value of integrated supply chain management. Glob. Supply Chain Technol. Manag. **1**(1), 216–225 (1998)