InterventionViz: Visual Analysis of Behavior-Change Intervention Dynamics

Category: Research

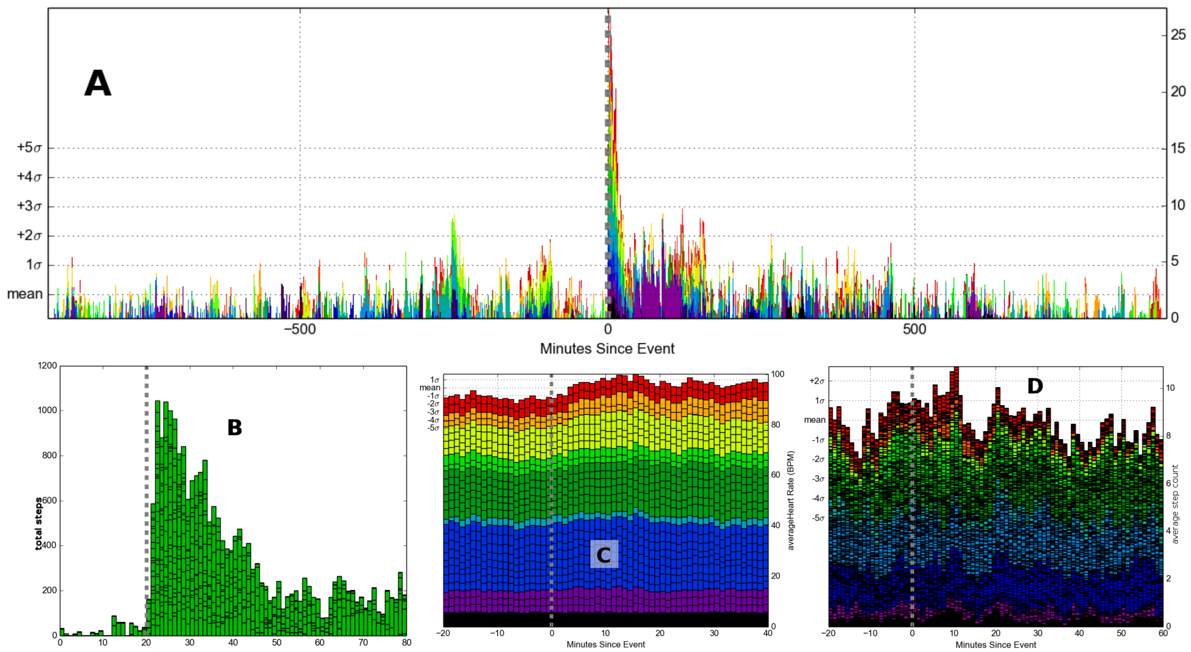

Fig. 1. A)Aggregated step counts surrounding a set of control intervention events shows event response dynamics and individual variations across events. B) Aggregation of step counts showing dramatic response to the the control intervention. C) Heart rate data aggregated across KNOWME participants shows a mild response to an SMS intervention. D) Aggregated step counts of subjects briefly exposed to an active mAvatar shows little trend.

Abstract— With the advent of research-grade wearable sensor suites and increasingly ubiquitous human-computer interface comes the opportunity for a new generation of behavioral theory and behavior change methodologies. Just-in-time adaptive interventions (JITAIs) designed to optimize a subjects behavior based on their momentary context are under exploration as a potentially powerful ally for practitioners and commercial behavior health applications. Analysis of the effect an intervention has on a target behavior, however, is a complex task not well handled by current methods.

In this work we explore the application of aggregated time-series visualization and system identification techniques to aid analysis of intervention event effects with emphasis on the dynamics of participant response to intervention. To highlight the strengths and caveats of the techniques employed, data from two trial studies of physical activity interventions (n=11, n=10) and one empirical control dataset (n=1) are utilized. The insights presented in this work offer a foundation for future research addressing this problem and as a guide for behavioral engineers.

Index Terms—Visualization in Social and Information Sciences, Hypothesis Testing, Visual Evidence, Time Series Data, Qualitative Evaluation, Biomedical and Medical Visualization

diamondrule-eps-converted-to.pdf

1 Introduction

Health sensing, machine learning, network access, and computation are rapidly becoming ubiquitous, enabling new tools for capturing human activities in the natural environment. Previous work shows such tools can be used to detect behaviors and psychological states such as stress[**?**, **?**], physical activity[**?**, **?**], social interaction[**?**], and smoking[**?**]. Likewise, there are many commercial, inexpensive, consumer-grade products (Jawbone Up, Fitbit, Nike Plus to name a few) that are putting these data in the hands of the layperson. The challenge that both laypeople and health research communities face is in making sense of this data.

Health researchers - and by extension overall public health - could benefit significantly from access to tools that leverage these data. Many of today's greatest health challenges can be mapped back to behavioral choices made in-the-moment. Avoiding physical activity, eating high-fat foods, and smoking cigarettes are all poor behavioral choices made in-the-moment that, in aggreggate, lead to a variety of health problems (e.g., obesity, diabetes, heart disease, cancer, chronic pain, depression), lower quality of life and shortened lifespans [?, ?, ?, ?]. Similar challenges exist outside the realm of personal health. Academic success is a function of attending class and completing assigned tasks, among other daily behaviors[?]. In personal finance, poor day-to-day purchasing decisions can add up to large financial debts[?].

There is a call in for the human-computer interaction (HCI), behavioral science, and other related communities to develop user interfaces for mobile behavioral interventions that help users make better in the moment behavioral choices related to health[?, ?], productivity[?, ?, ?], and more.

With the right tools, health researchers could better examine important behavioral events and coexisting contextual factors at a level of detail and richness not previously possible. Likewise, they can go beyond observing events. Behavioral researchers could deliver interventions to individuals in-the-moment, and then see how the intervention in turn changes behavior and affect.

In addition to metrics of success of an intervention, behavioral theorists need tools to help understand the dynamics of behavioral responses to a stimulus. It is due to the lack of a dynamical treatment of behavior within theories that existing models behavior appear inadequate to inform state-of-the-art intervention development [?]. Applicable methods of intervention analysis and data visualization have been slow to reach behavioral researchers, dramatically limiting their ability to develop of state-of-the-art behavioral theories to address these shortcomings. Unfortunately, the tools health researchers need to manage, review, and learn from such data do not yet exist. New methods for evaluating these behavioral interventions remain underexplored and conventional methods of analysis do not offer the level of detail needed to explore the implicit dynamics of just-in-time, interactive, or adaptive interventions.

In this paper methods for analysis of the dynamical response to an intervention are outlined. The impulse reponse of a physical activity intervention is explored and data visualizations which provide insigt into the dynamics of health-related events are demonstrated.

2 Previous Work

2.1 Related Event-based Time Series Visualization

"Lifelines" [?] allow for the exploration of events in a series for one individual, and new research in event sequence analysis [?], including analysis of event patterns [?, ?, ?] and the relation of multiple symptoms [?] helps researchers examine outcomes on a "macro-scale" across many subjects by aggregating records into a single view. Additionally, the problem of identifying patterns at multiple time scales has been partially addressed through clustering of time series [?], and methods for exploring the "paths" traversed by many individuals between many event types and statistical analyses to highlight relationships between events has recently been established [?].

Little existing work addresses the dynamics of a numerical variable's response to a behavioral intervention event. Statistical methods of intervention analysis [?], have been applied across various disciplines but thus far there has been little demand for these methods in behavioral science. Only recently has new wearable sensing technology made time-intensive, in-the-wild measurements feasible. Additionally, the prospect of ubiquitous intervention delivery via mobile devices and the concept of Just-in-Time-Adaptive-Interventions (JiTAIs) have introduced a new demand for a more detailed understanding of human behavior.

2.2 Current Intervention Methods

Recent advances in sensing and ubiquitous computing have enabled examination of and influence over behavior at relatively high frequency (on the order of seconds) and in a wide range of daily life contexts. New wearable sensing technologies are changing the way researchers do experiments, and mobile phones are a powerful new medium for delivering behavioral interventions "just-in-time".

Recent works have explored "adaptive interventions" tailored based on "tailoring variables" which may include user preferences, context, and personality [?]. Despite these drastic changes in the interventions, methods for evaluating adaptive interventions remain in many ways similar to "fixed" interventions [?], and the use of the multiphase optimization strategy (MOST) and sequential multiple assignment randomized trials (SMART) [?] maximize efficiency in applying these methods. These methods become more difficult to apply, however, when dealing with Just in Time Adaptive Interventions (JiTAIs) [?].

Existing work on visual analysis of systems usability [?] may be applied to the evaluation of JiTAI systems, however, these methods focus largely on a single record, rather than generalizations drawn from across many. Additionally, little theoretical guidance in terms of effect latency or delay exists to aid in the planning or analysis of experimental trials.

A theoretical basis which takes dynamical effects into consideration to enable improved behavioral intervention optimization has been proposed for interventions mediating gestational weight gain [?], smoking behavior [?], childhood anxiety [?], and fibromyalgia [?]. This new type of theoretical model is most effective on the timescale of multiple days, weeks, or months - partially because the confounds of contextual, within-day variations make analysis at this level difficult, but mostly because theories of behavior at this time scale are underdeveloped. Methods for analyzing the dynamics of intervention responses using existing data are needed in order to catalyze the development of theories to explain these signals.

3 Design Considerations

3.1 Characteristics of Intervention Datasets

Behavioral scientists with data-overload are becoming increasingly common as wearable sensors increase in popularity. There no doubt exist many under-utilized datasets with novel contributions to theory waiting to be discovered.

Common features of contemporary behavioral research dataset include:

- Multiple time-scales - multi-scale measurement at very different frequencies
- crossover designs - within-subject comparison is the preferred method for gauging efficacy of an intervention
- high-frequency numerical measures - Accelerometers, ECG, GPS, and much more
- numerical measures with low frequency - Ecological Momentary Assessment (EMA) constructs, blood-draws
- contextual, nominal data at various frequencies - activity classification, location classification, social contexts

4 Example Application: Physical Activity

As an example application to demonstrate the strengths of the proposed visual analytics, two empirical datasets will be used, each with a minute-level metric of physical activity and intervention events delivered throughout a period of several days. In both studies interventions were delivered with the intent of increasing subjects physical activity, and responses to interventions varied between participants and delivery contexts. In addition to this data, a control dataset with known intervention responses is included for comparison.

These datasets provide a good test bed for application of the methods presented here. Measurement of physical activity (PA) is a well-studied topic and many interventions focus on increasing physical activity, making PA a prime target for testing our methods. At the same time, the cognitive processes surrounding physical activity are familiar to most researchers and numerical representation of PA is easily interpreted.

The differences in the chosen datasets serve to highlight the strengths and weaknesses of methodologies outlined. The n-of-one control dataset with a strong intervention acts a baseline with predetermined response characteristics which should be easily identified by our analysis. The mAvatar study data shows less prominent effects

Data Set	n	length (days)	intervention	measures
control	1	14	N/A	step count
mAvatar	11	8+	glanceable avatar display	step count
KNOWME	10	3	SMS message	HR, accelerometry

Table 1. Summary of data sets used.

study wide, but has potentially interesting subgroups for exploration. Additionally the mAvatar data is unique in that it contains two interventions targeting the same theory, but influencing in opposing directions. Lastly, the KNOWME data represents a JiTAI with a study-wide effect and multiple behavioral measures.

The interventions in these datasets are all expected to primarily effect the level of the target behavior (see figure 3), but the dynamics of the response may differ greatly. The control intervention (by design) is expected to have minimal delay and a decay which starts 5m following the intervention. Thus the control intervention should closely resemble figure 4 (bottom) when viewed at the appropriate time scale. The mAvatar and KNOWME datasets each target very different psychological concepts which might be expected to have unique dynamic signatures.

4.1 Control Dataset

The control dataset is the result of manual recording of one subject undergoing an imaginary, very potent intervention. The subject remained sedentary for an interval ranging from 10 to 120 minutes. Then the subject was physically active for a period of no less than 5 minutes. A Fitbit One electronic pedometer was used to collect step counts as a proxy for physical activity.

4.2 mAvatar Study

An alternating treatment design is used to examine subject behavior over a period of 8+ days in order to test the effect size of an avatar-based live wallpaper deployed on Android phones [**?**]. Subjects (n=11) aged 11-14 were exposed to a simple, animated cartoon avatar on their mobile device showing alternating levels of physical activity. Each day the avatar would either be active (playing basketball, running, bicycling) or sedentary (watching TV, on a computer, or playing video games). Fitbit One electronic pedometers were used to estimate subject levels of physical activity via step count.

4.3 KNOWME Study

In this study ten teenagers were asked to carry a smartphone and wear an accelerometer and a heart rate monitor for 3 days. Physical activity was measured continuously and was monitored in real time using the KNOWME system [**?**]. When a subject had been continuously sedentary for two hours, a personalized SMS text message was sent to their phone. Each text message is manually crafted to prompt the subject to be more physically active.

5 Tasks

5.1 Highlighting Event Dynamics

In order to explore the dynamical response of an intervention, the shape of the input signal must be defined. In most cases an intervention can be represented as an impulse signal. Using this representation the impulse response can be calculated as the cross-correlation between the intervention signal and the behavioral measure. Figure 2 shows the result of cross-corrlation between the intervention input and the heart rate across all participants in the KNOWME study.

The dynamics surrounding a particular event can also be shown using a raw time series. The instance or span of the event is marked on the time-axis and the value of the behavioral measure (physical activity in this case) is encoded in the height at each point in time.

Figure 3 shows the case where an event instantaneously causes permanent change in the target behavior, but in the many cases the intervention will have a temporary effect on the target behavior and will have some delay before setting in.

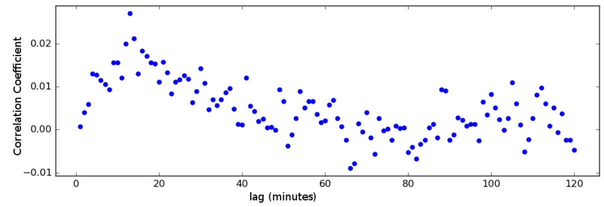

Fig. 2. CCF showing study-wide heart rate response to intervention for the KNOWME dataset.

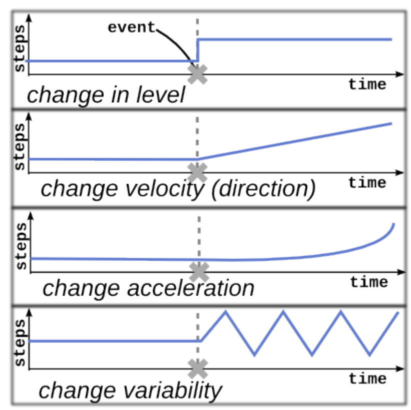

Fig. 3. Theoretical responses to an intervention (adapted from [**?**]).

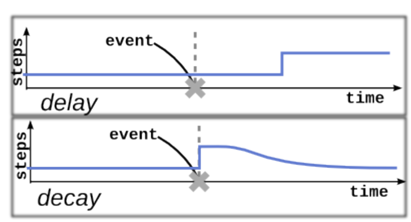

Fig. 4. Level-change responses with delay and decay (adapted from [**?**]).

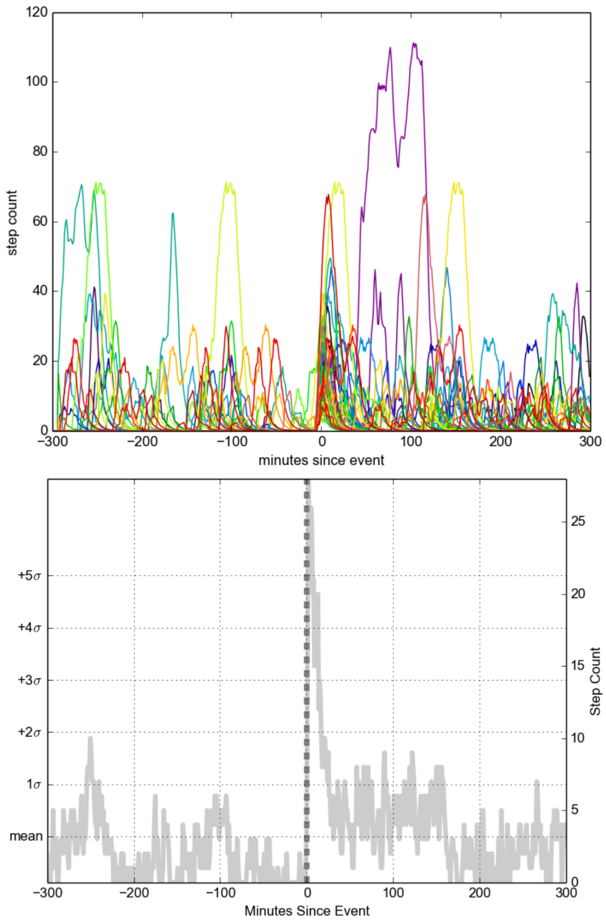

Fig. 5. Comparison of all event responses smoothed over a 15m rolling window (top) to average response (bottom) surrounding the control dataset intervention.

These intervention response dynamics (shown in figure 4) are critically important for just-in-time adaptive intervention developers, but are largely unaddressed in current theory. One method for examining event dynamics is to explore each participant's record individually (perhaps segmented by day), and to mark the events and manually explore individual responses. Since this examination is taking place over many series, it is prudent to utilize sparklines [**?**] or horizongraphs [**?**] to allow for examination of many series simultaneously. In doing this, however, it quickly becomes apparent that behavioral intervention data in this format is much too unwieldy. Random contextual fluctuations and inconsistency in frequency of intervention delivery makes visual identification of patterns extremely difficult when viewing an entire series, let alone series for multiple participants. Thus, we focus first on the characterizing the response to single events only, and can later address the issue of event history treated as a contextual subgrouping of all events.

5.2 Event-time Alignment

Plotting individual events one-by-one allows a researcher to explore the ideographic details of that particular event, but in order to draw out generalizations across groups of events (be it by participant, context, or any other selector) events must be plotted relative to the time of the event, rather than the start of the study. By time-shifting our data view so that each intervention event falls at t=0 in a time-series, we can view many events on a common time frame.

Figures 3 and 4 give a sense of what an intervention should look like, but in reality individual variations in context completely mask the often small effect of an intervention (see Figure 5 (top)). To a researcher looking at the plot of individual event responses in figure 5 (top), it might seem that only the intervention plotted in purple was an effective intervention, acting with a delay of approximately 30m, and decaying rapidly 120m after the event. However, the control dataset - by design - includes interventions that were 100% effective, acting with minimal delay and beginning decay at 5m. Plotting these series with aforementioned "summary dashboard" methods may allow researchers to identify this pattern more easily, but the average signal is still difficult to pull out of the noise. Luckily, this is a familiar problem with a familiar solution. Since our data has been time-shifted to place the time of event at t=0, an average of the series together can reveal nomothetic trends across all events. When looking at all events individually (figure 5 top), it is difficult to spot any pattern in the series. When averaging across all event responses, however, a clear, significant response is evident (figure 5 bottom), and the purple series is exposed as an outlier.

This approach can be taken for all events in one subject's time series to characterize that subject, or can be applied across subjects to characterize a more generalized response to the intervention. In fact, a subset of groups can even be selected and analyzed in order to enable advanced subgroup analysis.

As expected, figure 5 shows the control intervention to be quite effective at increasing the step count. The additional y-axis showing the mean and standard deviation of the series is included to give an increased sense of the significance of this effect relative to data which may be out of frame. In addition to the nearly immediate response, a longer-lasting effect reaching out to approximately 180m after the event seems to be boosting step count, though the all-events view in figure 5 as well as the stacked-events display in figure 1B reveals that there are two outlier events which may be the sole cause. These finding show how, although the average line-graph makes spotting effects easy, outlying events or participants can skew the average. This danger can be sometimes mitigated through use of median in place of mean, but since step-count does not obey a normal distribution (0 values are almost always modal), that approach does not work well in this case.

5.3 Stacking

To address the shortcomings of the aforementioned average-line shown in figure 5, all individual events can be shown stacked on a single graph. This aggregation method yields the same shape, and the y-axis can be easily normalized to match the average series by dividing by the number of events. While still evening out random contextual influences, this visual also provides indication that the average result is not due to one outlier event, enables easy spotting of missing data or faulty sensors, and gives some indication of the number of events considered. For an n-of-one dataset such as the control dataset, events can be graphed with a unique color. In figure 1A, event colors are chosen based on the order in which they were observed. Color mapping of events can also be used to visually group events based on time of day, location of the event, or participant.

For a plot of many participants, encoding participant in color allows the visual to display both event-level and event-group-level detail in addition to the overarching response. Figure 6 shows the difference between a plot of various average response lines and the stacked area plot using data from the KNOWME dataset. The thin lines in figure 6 (top) represent the response of each participant to the event averaged across all events for that participant. The thick gray line shows the average across all participants' average series. The stacked bars in figure 6 (bottom) are colored by participant ID, and each bar represents one unique event - stacked in order of event incidence. This allows researchers to search for both participant outliers within the set as well as event outliers within each participant. For instance, it is clear that the participant shown in purple responded to the intervention, due to the purple "bulge" but we can also see that this effect is largely the result of a single event within the participant's series. This reveals that the intervention was effective on average, while also showing that there exists some variable within participants moderating the efficacy of the intervention.

Figure 6 shows an increase in accelerometry counts following the delivery of a physical-activity-suggesting sms message. Though the

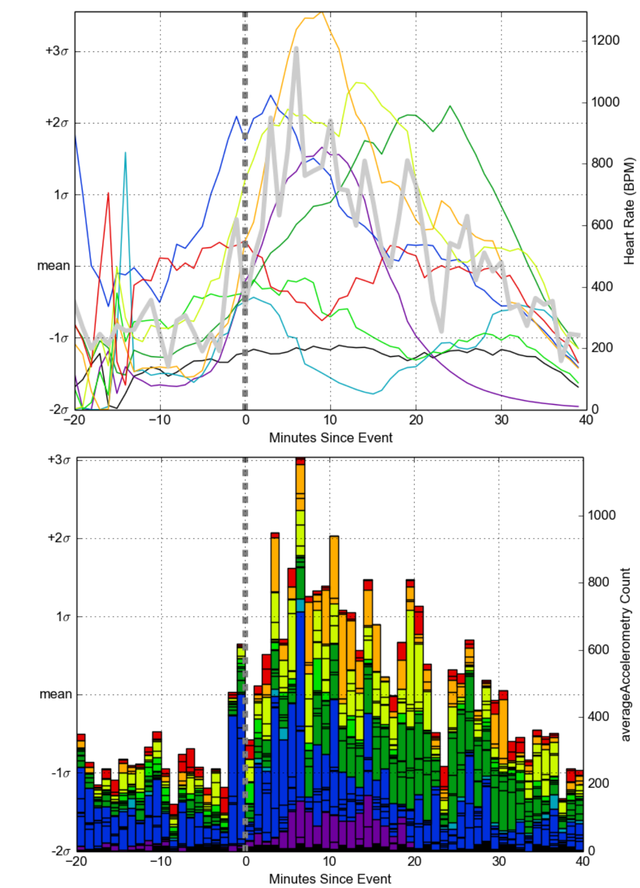

Fig. 6. Comparison of KNOWME average lines (smoothed over 15m rolling window) and stacked bars.

behavioral measure differs greatly from that used in the control dataset and 5, a comparison of the y-values in terms of standard deviation also reveals that this effect is less extreme than what we observe in the control intervention. The deviation from the mean as measured relative to the standard deviation gives a sense of how unlikely the signal is to be a random artifact, but detailed methods for evaluating the statistical likelihood of observing a particular shape are not covered here. For additional comparison to the control data, also consider the average-line view in figure 5, and the stackplot shown in figure 1B. Though the highlighted windows are relatively small (to highlight the intervention response), much wider context around the event can be plotted, such as that shown in figure 1A.

This same analysis is applied to figure 1C, but with another variable in the KNOWME dataset, heart rate. Both the accelerometry counts and heart rate signals should act as proxies of physical activity. Note however, the different dynamics of each variable's response. Accelerometry counts are more directly tied to behavior - which can be erratic and non-linear, thus the dynamics observed are more volatile, while heart rate acts as smoothed function of accelerometry, responding less quickly and decaying more slowly than accelerometry data.

The linegraph allows for characterization of unique individuals, but the stackplot better highlights the overall effect and also shows the number of events considered. Another key difference between these two approaches is the proportionate weighting of subjects into the global effect display. The line-graph approach considers each subject equally regardless of the number of events recorded of that subject, whereas the stackplot considers the events equally and thus each subject's contribution is weighted by the number of events recorded.

$$\bar{y}(t) = \frac{1}{N}\sum_{i=1}^{N} y_i(t) \tag{1}$$

$$\bar{y}(t) = \frac{1}{P}\sum_{p=1}^{P}\frac{1}{n_p}\sum i = 1_p^n y_{i,p}(t) \tag{2}$$

Equations (1) and (2) shows the difference in aggregation methods for line vs stacked views where N is the total number of events across all participants, P is the number of participants, n_p represents the number of events for subject p, y_i represents the time series for event i, and $y_{i,p}$ represents the time series for subject p's event i. Aggretating data via method (2) does a better job to ensure that one participant does not skew results, but can give too much weight to data from a participant with few events. In this particular case, it makes little difference, however, since the number of events per participant are roughly equal.

5.4 Characterize Intervention Delivery Context

In some cases introducing a control event against which to compare the experimental event can help isolate the intervention from the context in which it is delivered. For instance, an intervention delivered on a mobile device is always delivered within the context of phone interaction. That is, the user is always using the phone when the intervention is delivered. It is possible that "using the phone" has it's own unique effect on the behavioral measure. Thus, using "phone use" events as a baseline against which to compare "phone use and intervention delivery" strengthens the chance that the observed effect is a result of the intervention itself and not the result of frequently concurrent contextual forces. For example, by looking at all times the phone was viewed in the mAvatar dataset, the average context of phone use can be characterized.

In figure 7, a notable increase in steps leading up to phone usage is observed. It is possible that this increase - though it preempts avatar viewing - is indeed caused by the avatar. Consider, for instance, the unanimously reported case of subjects viewing the phone with the explicit purpose of seeing how the avatar would be affected by their behavior. Thus a peak in PA may indeed be driven by the desire to illicit a response from the avatar, which is viewed only a few minutes later. This interpretation is quite speculative and other features of figure 7 are not so easily explained. It is clear, however, that this is not a flat

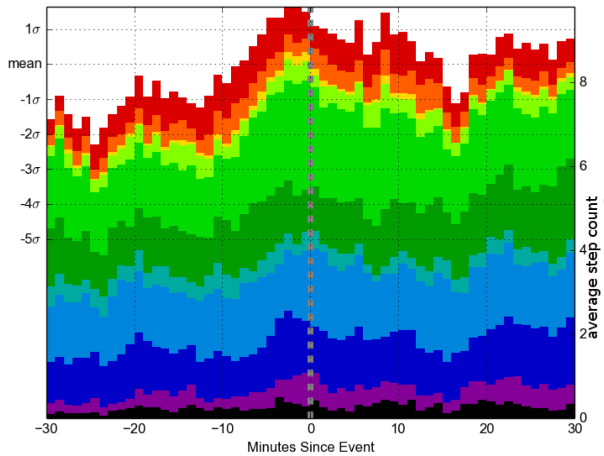

Fig. 7. Stackplot of step counts aggregates in the 30 minutes surrounding 1673 phone-view events from the mAvatar dataset (individual event segmentation removed due to large number of events).

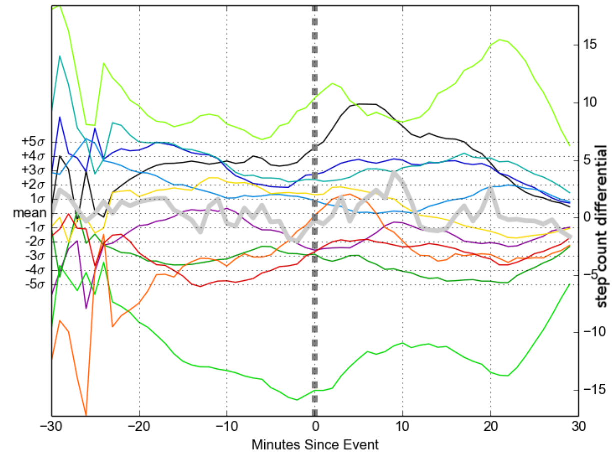

Fig. 8. Active-event series average minus sedentary-event series average smoothed over a 15m rolling window. (average across participants shown in bold)

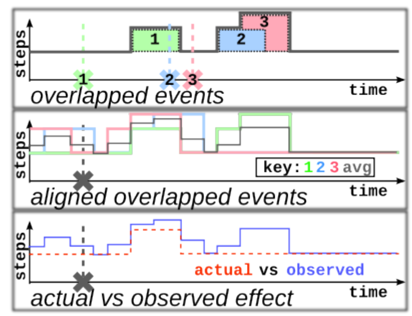

Fig. 9. Overlapping analysis windows confounds observed effect after time-alignment.

baseline that we may expect to find on average, and exploration of dynamics surrounding the active and sedentary avatar viewings ought to subtract this baseline to account for the overlapping of this context-driven (rather than event-driven) signal. This masking baseline signal is apparent in the unadjusted plot of active-avatar views shown in figure 1D. Rather than a comparison of each event vs the baseline, however, the specific case of the mAvatar dataset allows for a direct comparison between similar, opposing events (active-avatar vs sedentary-avatar).

5.5 Comparing Event Types

Aforementioned methods used to provide a contextual baseline of comparison for events can also be applied to allow for a comparison between two event types. By treating one event as the baseline, differences between the events can be visualized. Using this paradigm, nearly equivalent event responses will have a near-zero difference. Positively-valued areas of the resulting chart indicate times when the "experimental event" had a greater positive effect on the target measure, or, conversely, that the "control event" had a greater negative effect on the target measure.

The mAvatar dataset contains two types of intervention which may be interesting to compare: 1) active-avatar viewing, 2) sedentary-avatar viewing. In this case, the two event types are theoretically opposite in effect, meaning that the sedentary-avatar effect should resemble a mirrored version of the active-avatar effect. Thus, the difference should accentuate the intervention's effect signature and better isolate the behavioral response from noisy data.

Even with two oppositely-polarized events, however, figure 8 fails to show the dramatic effect a researcher might hope for. In this case, study investigators attribute the apparent lack of effect to an ambiguity in study design which led to two opposing conditions: 1) subjects respond positively to physically-active avatars via the Proteus Effect [**?**] 2) subjects respond negatively to physically-active avatars via falsely perceived biofeedback, and figure 8 may indeed suggest this subgrouping within the data in the individual participant series.

6 Discussion and Future Work

Though the presented work helps address some of the challenges faced by contemporary behavioral researchers, in some places there remains uncertainty in the meaning of the visualizations and even more deeply hidden discoveries. Additionally, new scientific questions have been raised through application of these visualizations and thus future work is required.

6.1 Dealing With Overlapped Data Frames

When looking at data surrounding an event, one must be cognizant of how instances of the same event at another time may effect our data. For instance, if an analysis targets the 30m following an event and the event frequently occurs at 10m intervals, the overlapping signals will create unwanted artifacts.

Figure 9 illustrates this point by showing how events falling within each other's windows of analysis confound the data and ultimately yields a signal quite unlike the actual effect response. In real data, this is further complicated by the non-linear way in which effects are expected to combine. An example of this effect in real data can be seen in the four yellow to lime green series plotted in figure 5 (top). The yellow peak near 0m represents a true intervention response, whereas the other three are artifacts introduced from relative nearness in time to the true response. In other words, the events analyzed in each of these series fall at approximately -280, -150, 0, and 120 minutes relative to the third event, and those artifacts all represent the same data. Study designs utilizing methods outlined in this paper should design studies to minimize overlapping analysis windows.

Event overlap becomes somewhat inevitable, however, for large event window sizes. Figure 10 shows a selection of data identical to that in figure 7, but without the inclusion of overlapping windows of analysis surrounding the events. Allowing no overlap between events helps ensure that multiple interventions effects do not skew the data, but ignoring these data points can drastically reduce the sample if large times following the event are used, because very few events are so isolated.

As is shown in 11, increasing the window of analysis to 12hours surrounding the phone-view event leaves only 46 events, and a noticeable increase in the variability of the data is observed.

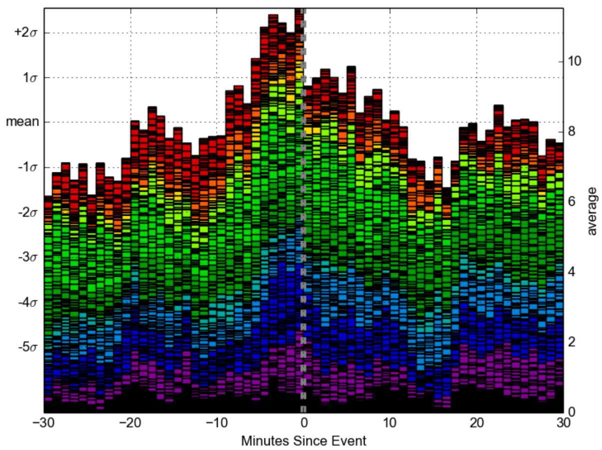

Fig. 10. Stackplot of step counts in the 30 minutes surrounding 586 phone-view events from the mAvatar dataset with no other events within 30min.

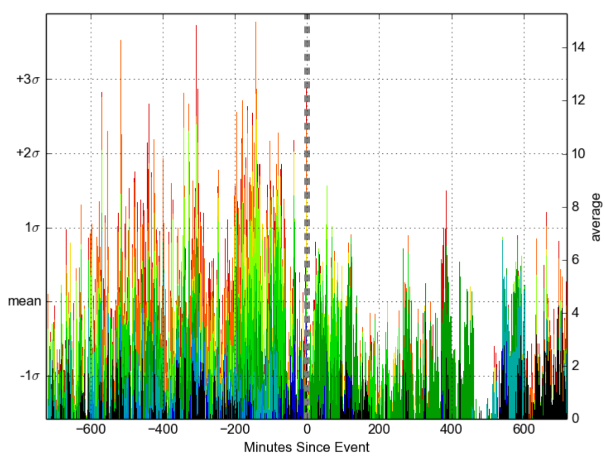

Fig. 11. Stackplot of step counts in the 12hrs surrounding 46 phone-view events from the mAvatar dataset with no other events within 12 hours.

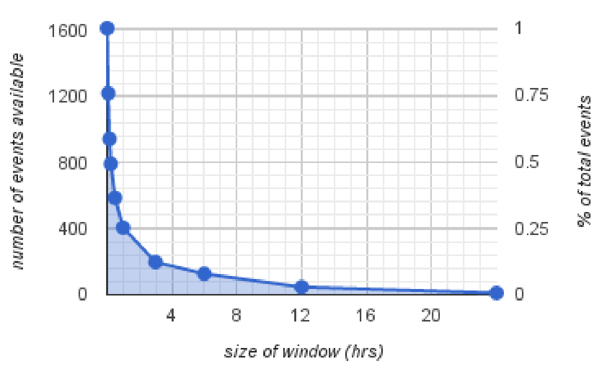

Fig. 12. Percent coverage of events in the mAvatar datset vs size of exclusion window surrounding the event.

For phone-view events among the population analyzed by the mAvatar study, we can estimate percent coverage of events available for "clean", non-overlapping analysis through the distribution shown in figure 12.

6.2 Alternative Stacked-Area Representation

Use of the "themeRiver/streamgraph" [**?**, **?**] paradigm for plotting stacked area charts may offer an improved view of the contribution of individual time series to the aggregated result, further easing the identification of outlier participants or events.

With better focus on individuals or subgroups, however, comes a reduced ability to look at the bigger picture. Thus, although a streamgraph may make for a better general view, the stacked area plots presented are still of use for those whose primary focus it is to evaluate the group response.

6.3 Characterizing Psychological Influence of Events via Response Signature

Different psychological mechanisms act on different time-scales and, likewise, with different dynamics. The delay of effect onset and decay of the effect observed in data could theoretically be used to suggest what psychological mechanisms are at work. In this way, interventions could be characterized in terms of applicable theory via the the dynamics observed.

This method becomes even more powerful when responses across multiple variables are considered. To draw an example from previously presented data, a combined analysis of heart rate (figure 1C) and accelerometer count (figure 6) dynamics improves the ability to match signals to known responses.

6.4 Statistical Analysis of Features

Much existing work on the statistical testing of between-phase differences in traditional AB study designs [**?**] is applied in the evaluation of the efficacy of a one-time or repeatedly applied intervention, and methods for evaluating the likelihood of features in a time-series are also well documented [**?**, **?**]. Through combination of existing intervention analysis techniques [**?**], goodness-of-fit evaluations of model formulations [**?**] in comparison to surrogate time series, and the presented visualization methods, researchers have a good foundation for analyzing dynamic models of human behaviors.

7 Conclusion

There are many reasons why behavior change is hard: people build up habits over time, behaviors may be tied to addictive substances or activities, behavior is tied to normal daily human experience and context (i.e. smoking more often when with a certain group of individuals or in a particular location). The full potential of behavioral interventions - particularly JITAIs - will continue unrealized until mechanistic methods of behavior modeling are adopted by the behavior science community.

The presented visualization methods allow for more detailed analysis of how a subjects behavior responds to a stimuli over time. These methods, when combined with a computational modeling approach to understanding human behavior may enable behavioral scientists to formulate more accurate and more application-ready theories, leading to more effective behavioral interventions.

References

[1] G. E. Box and G. C. Tiao. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical association*, 70(349):70–79, 1975.

[2] L. Byron and M. Wattenberg. Stacked graphs–geometry & aesthetics. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1245–1252, 2008.

[3] K.-h. Chang, D. Fisher, J. Canny, and B. Hartmann. How's my mood and stress?: an efficient speech analysis library for unobtrusive monitoring on mobile phones. In *Proceedings of the 6th International Conference on Body Area Networks*, pages 71–77. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011.

[4] L. M. Collins, S. A. Murphy, and K. L. Bierman. A conceptual framework for adaptive preventive interventions. *Prevention science*, 5(3):185–196, 2004.

[5] L. M. Collins, S. A. Murphy, and V. Strecher. The multiphase optimization strategy (most) and the sequential multiple assignment randomized trial (smart): new methods for more potent ehealth interventions. *American journal of preventive medicine*, 32(5):S112–S118, 2007.

[6] H. Cooper, J. C. Robinson, and E. A. Patall. Does homework improve academic achievement? a synthesis of research, 1987–2003. *Review of educational research*, 76(1):1–62, 2006.

[7] S. Deshpande, D. E. Rivera, J. W. Younger, and N. N. Nandola. A control systems engineering approach for adaptive behavioral interventions: illustration with a fibromyalgia intervention. *Translational behavioral medicine*, 4(3):275–289, 2014.

[8] Y. Dong, D. E. Rivera, D. S. Downs, J. S. Savage, D. M. Thomas, and L. M. Collins. Hybrid model predictive control for optimizing gestational weight gain behavioral interventions. In *American Control Conference (ACC), 2013*, pages 1970–1975. IEEE, 2013.

[9] A. L. Dunn, M. H. Trivedi, and H. A. O'Neal. Physical activity dose–response effects on outcomes of depression and anxiety. *Medicine & Science in Sports & Exercise*, 2001.

[10] B. A. Emken, M. Li, G. Thatte, S. Lee, M. Annavaram, U. Mitra, S. Narayanan, and D. Spruijt-Metz. Recognition of physical activities in overweight hispanic youth using knowme networks. *Journal of physical activity & health*, 9(3):432, 2012.

[11] J. A. Fails, A. Karlson, L. Shahamat, and B. Shneiderman. A visual interface for multivariate temporal data: Finding patterns of events across multiple histories. In *Visual Analytics Science And Technology, 2006 IEEE Symposium On*, pages 167–174. IEEE, 2006.

[12] O. H. Franco, C. de Laet, A. Peeters, J. Jonker, J. Mackenbach, and W. Nusselder. Effects of physical activity on life expectancy with cardiovascular disease. *Archives of internal medicine*, 165(20):2355–2360, 2005.

[13] G. V. Glass, V. L. Willson, and J. M. Gottman. *Design and analysis of time-series experiments*, volume 197. Colorado Associated University Press Boulder, 1975.

[14] B. S. Gorman and D. B. Allison. Statistical alternatives for single-case designs. 1996.

[15] D. Gotz and H. Stavropoulos. Decisionflow: Visual analytics for high-dimensional temporal event sequence data. 2014.

[16] B. L. Harrison, R. Owen, and R. M. Baecker. Timelines: an interactive system for the collection and visualization of temporal data. In *Graphics Interface*, pages 141–141. Citeseer, 1994.

[17] S. Havre, B. Hetzler, and L. Nowell. Themeriver: Visualizing theme changes over time. In *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pages 115–123. IEEE, 2000.

[18] J. Ho and S. S. Intille. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 909–918. ACM, 2005.

[19] S. Jewell. Productivity via mobile phones: Using smartphones in smart ways. *Journal of Electronic Resources in Medical Libraries*, 8(1):81–86, 2011.

[20] P. Klasnja and W. Pratt. Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of biomedical informatics*, 45(1):184–198, 2012.

[21] M. Li, V. Rozgic, G. Thatte, S. Lee, B. Emken, M. Annavaram, U. Mitra, D. Spruijt-Metz, and S. Narayanan. Multimodal physical activity recognition by fusing temporal and cepstral information. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(4):369–380, 2010.

[22] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 351–360. ACM, 2012.

[23] U. Mitra, B. A. Emken, S. Lee, M. Li, V. Rozgic, G. Thatte, H. Vathsangam, D. Zois, M. Annavaram, S. Narayanan, et al. Knowme: a case study in wireless body area sensor network design. *Communications Magazine, IEEE*, 50(5):116–125, 2012.

[24] T. Murray, L. Jaimes, E. Hekler, D. Spruijt-Metz, and A. Raij. A glanceable mobile avatar for behavior change. In *Proceedings of the 4th Conference on Wireless Health*, page 16. ACM, 2013.

[25] I. Nahum-Shani, M. Qian, D. Almirall, W. E. Pelham, B. Gnagy, G. A. Fabiano, J. G. Waxmonsky, J. Yu, and S. A. Murphy. Q-learning: A data analysis method for constructing adaptive interventions. *Psychological methods*, 17(4):478, 2012.

[26] I. Nahum-Shani, S. N. Smith, A. Tewari, K. Witkiewitz, L. M. Collins, B. Spring, and S. Murphy. Just in time adaptive interventions (jitais): An organizing framework for ongoing health behavior support. *Methodology Center technical report*, (14-126), 2014.

[27] J. M. Norvilitis, P. B. Szablicki, and S. D. Wilson. Factors influencing levels of credit-card debt in college students1. *Journal of Applied Social Psychology*, 33(5):935–947, 2003.

[28] A. Pankratz. *Forecasting with dynamic regression models*, volume 935. John Wiley & Sons, 2012.

[29] R. I. Parker and D. F. Brossart. Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy*, 34(2):189–211, 2003.

[30] A. A. Pina, L. E. Holly, A. A. Zerr, and D. E. Rivera. A personalized and control systems engineering conceptual approach to target childhood anxiety in the contexts of cultural diversity. *Journal of Clinical Child & Adolescent Psychology*, 43(3):442–453, 2014.

[31] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: visualizing personal histories. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 221–227. ACM, 1996.

[32] H. Reijner et al. The development of the horizon graph. 2008.

[33] W. T. Riley, D. E. Rivera, A. A. Atienza, W. Nilsen, S. M. Allison, and R. Mermelstein. Health behavior models in the age of mobile interventions: are our theories up to the task? *Translational behavioral medicine*, 1(1):53–71, 2011.

[34] R. Ross, D. Dagnone, P. J. Jones, H. Smith, A. Paddags, R. Hudson, and I. Janssen. Reduction in obesity and related comorbid conditions after diet-induced weight loss or exercise-induced weight loss in mena randomized, controlled trial. *Annals of internal medicine*, 133(2):92–103, 2000.

[35] E. Sazonov, K. Metcalfe, P. Lopez-Meyer, and S. Tiffany. Rf hand gesture sensor for monitoring of cigarette smoking. In *Sensing Technology (ICST), 2011 Fifth International Conference on*, pages 426–430. IEEE, 2011.

[36] T. Sohn, K. A. Li, G. Lee, I. Smith, J. Scott, and W. G. Griswold. *Place-its: A study of location-based reminders on mobile phones*. Springer, 2005.

[37] H. K. Suen and D. Ary. *Analyzing quantitative behavioral observation data*. Psychology Press, 1989.

[38] K. P. Timms, D. E. Rivera, M. E. Piper, and L. M. Collins. A hybrid model predictive control strategy for optimizing a smoking cessation intervention. In *American Control Conference (ACC), 2014*, pages 2389–2394. IEEE, 2014.

[39] E. R. Tufte. Beautiful evidence. *New York*, 2006.

[40] J. J. Van Wijk and E. R. Van Selow. Cluster and calendar based visualization of time series data. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*, pages 4–9. IEEE, 1999.

[41] K. Vrotsou, K. Ellegard, and M. Cooper. Everyday life discoveries: Mining and visualizing activity patterns in social science diary data. In *Information Visualization, 2007. IV'07. 11th International Conference*, pages 130–138. IEEE, 2007.

[42] K. Wongsuphasawat and D. Gotz. Outflow: Visualizing patient flow by symptoms and outcome. In *IEEE VisWeek Workshop on Visual Analytics in Healthcare, Providence, Rhode Island, USA*, 2011.

[43] K. Wongsuphasawat and D. Gotz. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2659–2668, 2012.

[44] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman. Lifeflow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1747–1756. ACM, 2011.

[45] D. Wyatt, T. Choudhury, J. Bilmes, and J. A. Kitts. Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):7, 2011.

[46] D. G. Yanbaeva, M. A. Dentener, E. C. Creutzberg, G. Wesseling, and E. F. Wouters. Systemic effects of smoking. *Chest Journal*, 131(5):1557–1566, 2007.

[47] N. Yee, J. N. Bailenson, and N. Ducheneaut. The proteus effect. *Com-*

munication Research, 36(2):285–312, 2009.