Using Data Mining to Refine Digital Behaviour Change Interventions

Nathaniel Charlton
University of Brighton
Brighton, Sussex, United Kingdom
Do Something Different Ltd
Brighton, Sussex, United Kingdom
b.charlton@brighton.ac.uk

John Kingston
University of Brighton
Brighton, Sussex, United Kingdom
j.k.kingston@brighton.ac.uk

Miltos Petridis
Middlesex University
London, United Kingdom
m.petridis@mdx.ac.uk

Ben (C) Fletcher
University of Hertfordshire
Hertfordshire, United Kingdom
Do Something Different Ltd
Brighton, Sussex, United Kingdom
b.fletcher@herts.ac.uk

ABSTRACT

Do Something Different (DSD) behaviour change interventions are digitally delivered programmes designed to help people improve their health and wellbeing by adopting healthier habits. In addition to content addressing specific issues, such as diet, smoking and stress reduction, DSD interventions contain a core component promoting *behavioural flexibility*. This component helps people practice behaving in ways they currently do not, such as assertively, proactively or spontaneously, and is based on a model developed by psychologists researching the connections between behavioural flexibility and wellbeing.

This paper describes how we have used data mining techniques to optimise the design of DSD interventions, in particular the behavioural flexibility component. We present correlation networks and regression models obtained using pre- and post-intervention questionnaire data from 15,550 people who have participated in a DSD intervention delivered by email, SMS or smartphone app. We explain how these results led us to a clearer understanding of the connections between behaviour and wellbeing, using which we have optimised DSD interventions, ensuring that participants concentrate on developing the behaviours that are likely to benefit them the most.

Additionally we have used logistic regression to fit a propensity score model, which models how likely it is that each person in the dataset will complete the post-intervention questionnaire, based on their pre-intervention questionnaire data. When we stratify our dataset using these propensity scores, we find that the kind of people who are the least likely to tell us they have completed the intervention, by answering the post-intervention questionnaire, are also the kind of people who will experience the biggest increase in wellbeing from a completed programme.

CCS CONCEPTS

• **Information systems → Data mining**; • **Applied computing → Health informatics**; *Psychology*;

KEYWORDS

behaviour change; wellbeing; data mining; m-health

ACM Reference format:
Nathaniel Charlton, John Kingston, Miltos Petridis, and Ben (C) Fletcher. 2017. Using Data Mining to Refine Digital Behaviour Change Interventions. In *Proceedings of DH '17, London, United Kingdom, July 02-05, 2017,* 9 pages. https://doi.org/http://dx.doi.org/10.1145/3079452.3079468

ACM ISBN 978-1-4503-5249-9/17/07…$15.00
https://doi.org/http://dx.doi.org/10.1145/3079452.3079468

1 INTRODUCTION

Do Something Different (DSD) behaviour change interventions are digitally delivered (m-health) programmes designed to help people improve their health and wellbeing by adopting healthier habits. Since 2012, Do Something Different Ltd has designed and delivered a wide range of DSD interventions, addressing health and wellbeing issues such as stress reduction, weight loss, smoking cessation and diabetes self-management, as well as broader personal development objectives such as leadership. Results have been reported previously in a white paper [11].

Each DSD intervention begins with an online pre-programme questionnaire, where the user answers questions about their behaviours, habits, wellbeing, thoughts and feelings. Then, over the next few weeks, the participant receives a series of personalised recommendations of small activities, called "Dos", that are outside their normal habits [5]:

- On a smoking cessation programme, a user who has answered that they often smoke "while sitting in your favourite place/chair/spot on the sofa" might be advised "Today

break one connection: only smoke when standing up. Don't take one sitting down!".
- On a stress reduction programme, a user who answered that they rarely "feel positive" might be prompted, "Today write a list of things that have gone well for you lately. Even small things make life good, don't overlook the ordinary."
- On a happiness programme, a user who answered that they rarely "do things that make you feel good" might recieve the following "Do": "Today make time to do something which you know makes you feel good. Put music on, make your favourite meal, relax in the bath or recall your favourite memories."
- On a programme targeting physical inactivity, a user who answered that they often "spend much of the day in front of a screen" might be sent this "Do": "Take 15! Set a timer to turn off your TV/shut down your screen for 15-minute breaks. Use the time to walk outside, get some fresh air."

Further examples of "Dos" are given in [5]. By focusing on actions — doing — rather than just thinking, "Dos" are designed to bring about actual behaviour change rather than simply offering information; they are positive actions, small steps towards a bigger goal that are designed to be fun and motivating.

"Dos" are delivered digitally by smartphone app push notification, SMS message or email, and are supported by other material such as motivational messages and inspirational quotes. Participants also have access to a "Do Zone", an online community where they can share their experiences in a variety of forms and record their progress. Participants are offered the chance to complete the questionnaire again after their programme; doing so gives them access to a personalised report comparing their pre- and post-programme scores. Programmes have had an average duration of 6 weeks, and contained an average of 20 "Dos".

While many of the "Dos" in a DSD intervention are directly related to the objective of the programme, as in the examples above, a subset of the "Dos" on each intervention also aim to promote *behavioural flexibility*. This aspect of the interventions helps people practice behaving in ways they currently do not, or that are outside their comfort zone, such as behaving assertively, proactively or spontaneously. A person who answered that they do not behave assertively might receive a prompt, "Be a bit more assertive today: Speak up when you would normally hold back. Be direct in asking for what you want." A person who answered that they do not behave proactively may be advised: "Do something today to make tomorrow easier. Lay out your clothes, make tomorrow's lunch, fill up the car, empty your inbox. Enjoy proactivity." These "Dos" are known as *expanders*, as they aim to expand the person's range of behaviour. DSD's behaviour model is based on findings from a series of papers and books by psychologists Fletcher, Pine and others (e.g. [4–7, 10]), and includes 30 behaviours.

In order to make sure that future participants get as much benefit from their programme as possible, we wanted to understand how each of these 30 behaviours contributes to wellbeing. That way, programmes could be optimised to concentrate on helping people develop the behaviours that are likely to benefit them the most. Since completion of the post-intervention questionnaire is optional, we also wanted to understand any patterns in the kinds of users who are more or less likely to complete it. This paper describes how we have used data mining techniques to make progress on these issues, analysing questionnaire data for a sample of 15,550 people who have taken part in a DSD intervention.

Firm	Gentle
Unpredictable	Predictable
Individually-centred	Group-centred
Behave as you wish	Behave as others want you to
Reactive	Proactive
Lively	Not lively/Laid back
Calm/Relaxed	Energetic/Driven
Play it safe	Risk-taker
Conventional	Unconventional
Open-minded	Single-minded
Assertive	Unassertive
Introverted	Extroverted
Systematic	Spontaneous
Flexible	Definite
Trusting	Wary of others

Figure 1: The 30 behaviours included in the behaviour rater, organised into 15 pairs of opposites.

2 DATA SET USED

Our dataset consists of pre-intervention questionnaire responses from 15,550 people who participated in a DSD intervention, and post-intervention answers to the same questionnaire for 3,033 of these people. Here we examine two sections of the questionnaire:

Behaviour rater (full details in [5]) The participant is shown a 6×5 grid, each cell containing a description of a behaviour, and instructed: "Click on the behaviours below that best describe you. Select as many or as few as you like, so long as they describe how you generally are." The 30 behaviours consist of 15 pairs of opposites (positioned far apart in the grid), as shown in Figure 1.

Wellbeing questions Participants are shown 8 statements and asked, "Thinking about how your life has been in the last month, move each slider to indicate how much you agree with the wellbeing statements." Each person's 8 slider positions are converted to integers from 0 (the "a little" end) to 100 (the "a lot" end) and summed to give a wellbeing score from 0 to 800, higher values indicating better wellbeing. The questionnaire is similar to the Warwick–Edinburgh Mental Wellbeing Scale [15], addressing feeling and functioning aspects of wellbeing, e.g. finding it easy to make decisions or feeling happy. The questions show high internal consistency (Cronbach's $\alpha = 0.89$, pooling the pre- and post-intervention data).

Mean wellbeing scores are shown in the first column of Table 1. Here and elsewhere in the paper, we report results for four sets of data:

(1) The pre-intervention data for all participants.
(2) The pre-intervention data for just those participants who went on to complete the post-intervention questionnaire.

Table 1: Summary statistics of the data set. (All correlations shown are significant with two-tailed $p < 2 \times 10^{-11}$.)

	Wellbeing score (mean)	Beh. flexibility score (mean)	Correlation
pre-intervention (all users)	486.9	18.66	**0.18**
pre-intervention (users with post-data available)	510.0	20.24	**0.15**
post-intervention	561.2	19.49	**0.18**
increase from pre- to post-	51.1	−0.75	**0.12**

(3) The post-intervention data.
(4) The increases from pre-intervention to post-intervention; here we subtract the pre-intervention scores from the post-intervention scores. Thus positive values represent an increase over the course of the intervention, and negative values represent a decrease.

The positive change in mean wellbeing seen in Table 1 confirms that the interventions provide an improvement in wellbeing, as reported in [11].

3 EXISTING MODEL OF BEHAVIOUR AND WELLBEING

The existing DSD behaviour model (including the behaviour rater instrument) came from [5], which sets out the theory that behavioural flexibility can help explain the differences in wellbeing experienced by different people. According to this viewpoint, some people have a smaller range of behaviours to call upon to meet the challenges that arise in their lives and thus experience more stress and difficulty than others. By contrast, a flexible person is thought of as one who is able to behave in a wide range of ways. Instead of being solely extroverted or solely introverted, for example, a behaviourally flexible person can use either introverted behaviour or extroverted behaviour as the situation demands:

> "Consider for a moment the extrovert who is the life and soul of the party and happy being the centre of attention. His extroversion is not always an asset. In fact it becomes a handicap when he's forced to have a quiet night in, or on a visit to his girlfriend's sombre parents. The introvert on the other hand may cling to the walls at a wild party, but knows how to enjoy his own company or that of more serious folk. A person who can flex, using extroversion and introversion traits appropriately, is equally comfortable in either context. His personality does not alienate him from any corner of the world." [5]

The psychologists [5] propose a formula for scoring a person's answers on the behaviour rater (i.e. their set of selected behaviours), called the *behavioural flexibility score*:

$$100\% \times \frac{1}{2} \left(\begin{array}{c} \text{no. of behaviours selected} \times \frac{1}{30} \\ + \text{ no. of opposite pairs with both selected} \times \frac{1}{15} \end{array} \right) \quad (1)$$

Table 2: Correlations between measures of facilitatory / inhibitory behaviour and wellbeing score. (All correlations shown are significant with two-tailed $p < 2 \times 10^{-16}$.)

	Number of facilitatory behaviours	Number of inhibitory behaviours	No. of fac. beh's minus no. of inh. beh's
pre-intervention (all users)	**0.36**	**-0.33**	**0.45**
pre-intervention (users with post-data available)	**0.34**	**-0.33**	**0.45**
post-intervention	**0.35**	**-0.27**	**0.45**
increase from pre- to post-	**0.21**	**-0.15**	**0.27**

Higher scores indicate greater behavioural flexibility. Table 1 shows the means of the pre- and post-programme behavioural flexibility scores. The third column shows Pearson correlation coefficients between wellbeing scores and behavioural flexibility scores. We see that indeed higher behavioural flexibility scores are associated with higher wellbeing scores.

However, we also note that while wellbeing rises post-intervention, behavioural flexibility scores actually fall slightly. This suggested to us that the formula (1) could be improved upon in terms of capturing the relationship between behaviour and wellbeing. Specifically, we can draw out for investigation two hypotheses that are implicit in the formula (1):

- The formula is monotonic: adding an extra behaviour always increases the score. Thus, implicit in the formula is that none of the 30 behaviours are "bad for you"; adding a new behaviour to one's repertoire is *always* a good idea because it gives one an extra tool in one's toolbox with which to meet life's demands. For convenience we shall name this idea *Every Behaviour Is Useful* (EBIU).
- The formula also awards a boost in score when someone selects both of a pair of opposite behaviours, e.g. "extroverted" and "introverted"; rather than viewing this as contradictory, the model [5] interprets it as evidence of flexibility, in that the person has the capacity to be either extroverted or introverted as each situation demands. Let us call this idea *Opposites Are Special* (OAS).

4 RELATIONSHIPS BETWEEN THE BEHAVIOURS

We proceeded by constructing a correlation network. Correlation networks provide a way of visualising the key relationships among a large number of variables, and have been used previously [2, 3] for studying personality data. The idea is to take the variables (in this case behaviours) as nodes and put edges between pairs of variables that show (relatively) high correlation.

Specifically, we used the ϕ coefficient as our measure of correlation, on the full pre-intervention data, with a threshold of 0.175.

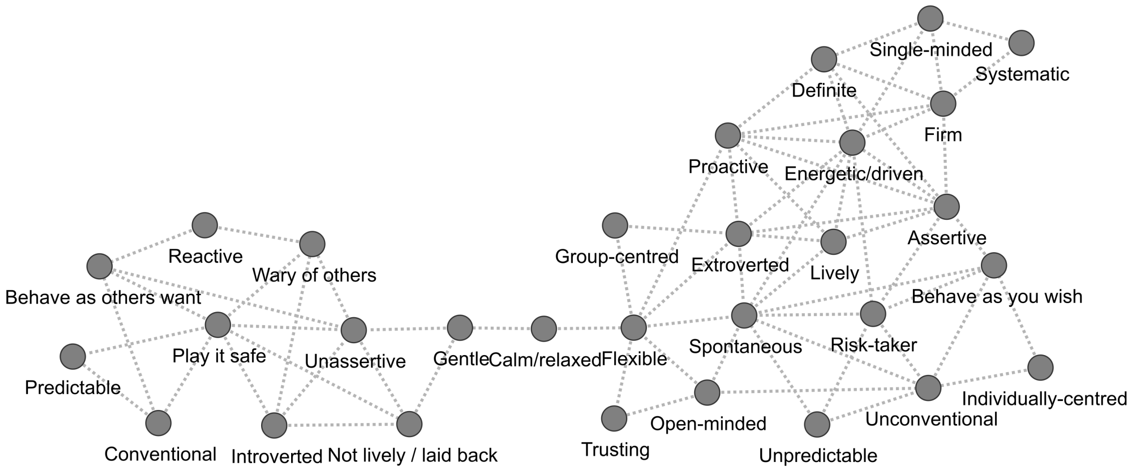

Figure 2: A correlation network depicting the typical co-occurrences of the 30 behaviours.

Table 3: Average number of facilitatory and inhibitory behaviours selected on the behaviour rater.

	No. of facilitatory behaviours	No. of inhibitory behaviours
pre-intervention (all users)	6.88	2.21
pre-intervention (users with post-data available)	7.35	2.26
post-intervention	7.43	1.72
increase from pre- to post-	0.08	−0.54

This left the two behaviours "reactive" and "individually-centred" as isolated nodes so, at our own discretion, we also added edges from each of these nodes to the two most correlated behaviours. This led to the final network show in Figure 2. The threshold of 0.175 was chosen by eye[1] to give an interesting but digestible network. Choosing a very low value for the threshold results in a network with an overwhelming number of edges and in which little structure is apparent; on the other hand, choosing a large value for the threshold results in a very sparse network where most of the nodes are isolated and, again, little structure is apparent.

The first thing we notice is that the behaviours have separated into two main groups, with three connecting "bridge" nodes: "gentle", "calm/relaxed" and "flexible". The behaviours to the left of the "bridge" appear to share a common theme: they generally appear to reflect a narrowing down of a person's options for action. If a person is wary of others, they are unlikely to take actions that others may disapprove of; while if a person is unassertive, they may be uncomfortable even stating what actions they wish to take. In either case, their possible behavioural options are restricted. We thus term these *inhibitory* behaviours. Conversely, many of the remaining behaviours (the "bridge" and to the right of it) appear to be linked to having a broader range of possible actions available in any situation. For example an open-minded person may see more options and an unconventional person may be less restricted by social conventions. We term these *facilitatory* behaviours. The data, therefore, suggest that a higher-order variable connects the 30 behaviours, even though they represent separate traits [5].

Table 2 shows the correlations between the wellbeing score and the number of facilitatory behaviours, the number of inhibitory behaviours and the difference between the two. These correlations are much stronger than those in Table 1 (in bold), indicating that the concept of facilitatory/inhibitory behaviours does a much better job of explaining wellbeing than the behavioural flexibility formula (1). Table 3 gives the average numbers of facilitatory and inhibitory behaviours selected on the pre- and post-intervention behaviour rater, and the average changes in these. Intervention participants on average lose 0.54 inhibitory behaviours and gain 0.08 facilitatory ones.

These findings suggest an alternative explanation of why DSD interventions work: the interventions help the participants to lose their inhibitory behaviours, which reduce their wellbeing, and also bring about a smaller increase in facilitatory behaviours, which increase their wellbeing.

5 REGRESSION MODELS LINKING BEHAVIOURS AND WELLBEING

In our research we have experimented with a range of regression models which explain wellbeing scores using behaviours. Here we report and compare results for the following (linear, ordinary least squares) models:

- *Behavioural flexibility model:* a model with a single predictor, the behavioural flexibility score from formula (1).
- *Behaviour count and opposites count model:* a model with two predictors, the number of behaviours selected and the

[1]We viewed the graphs for various thresholds using the Kamada–Kawai layout algorithm as implemented in the Visone program, available from http://visone.info/.

Table 4: Performance, with 10-fold cross-validation, of models predicting wellbeing scores from behaviours. Models are listed in order of RMSE on the full pre-intervention data.

Model	pre-intervention (all users)		pre-intervention (users with post- data)		post-intervention		increase from pre- to post-	
	RMSE	Correlation	RMSE	Correlation	RMSE	Correlation	RMSE	Correlation
Individual beh's & opposites count model	122.2	**0.56**	121.9	**0.55**	115.4	**0.53**	119.9	**0.28**
Individual behaviours model	122.2	**0.56**	121.8	**0.55**	115.4	**0.53**	120.0	**0.28**
Facilitatory/inhibitory model	130.9	**0.46**	129.3	**0.46**	120.2	**0.46**	120.3	**0.27**
Behaviour count & opposites count model	145.3	**0.17**	144.4	**0.14**	133.2	**0.19**	124.1	**0.11**
Behavioural flexibility model	145.3	**0.17**	144.3	**0.14**	133.6	**0.17**	124.1	**0.11**
Intercept-only model	147.6		145.8		135.6		124.9	

Table 5: Coefficients for the facilitatory/inhibitory model.

Predictor	Regression coefficients									
	Pre-intervention (all users)				Pre-int. (users with post- data)		Post-int.		Increase from pre- to post-	
	Mean	Min	Max	SD	Mean	SD	Mean	SD	Mean	SD
(Intercept)	443.9	442.6	445.7	0.95	467.2	2.73	503.5	2.06	43.6	0.65
No. of facilitatory beh's	13.5	13.3	13.7	0.13	12.7	0.25	12.9	0.21	8.7	0.23
No. of inhibitory beh's	−22.5	−22.8	−22.2	0.17	−22.5	0.50	−22.4	0.24	−12.7	0.53

number of opposite pairs where both were selected. (This is like the previous model, except that the two terms from the formula (1) have their coefficients fitted separately.)

- *Facilitatory/inhibitory model*: a model with two predictors, the number of facilitatory behaviours selected and the number of inhibitory behaviours selected.
- *Individual behaviours model*: a model with 30 binary predictors, one for each behaviour in the behaviour rater[2].
- *Individual behaviours and opposites count model*: this is the individual behaviours model, extended with an extra integer-valued predictor, namely the number of opposite pairs of behaviours where both were selected.

Our baseline for comparison is an *intercept-only* model, which simply always predicts the mean wellbeing score in the dataset.

We evaluated and compared our models using 10-fold cross-validation, as widely practiced and recommended (see e.g. [8, §7.10]). Cross-validation gives a way to assess how well a predictive model is likely to perform on unseen data, and guards against overfitting. For each of the four sets of data Table 4 reports two measures of model performance: RMSE and correlation coefficients. The reported RMSE (root mean squared error) values are the means of the RMSEs for each of the 10 folds. The correlations reported are Pearson correlation coefficients obtained using a *pooling* strategy, i.e. we bring together the pairs of actual and predicted scores from the 10 folds, and calculate the correlation on this combined set of points.

The relative performance of the models is very consistent across the four sets of data. In terms of predictive power, the existing behavioural flexibility model is not much better than the intercept-only model, and nor is the behaviour count and opposites count model.

The facilitatory/inhibitory model is significantly better, indicating that our division of behaviours into two groups, from Section 4, does provide a useful tool for understanding behaviours and their effects. The fitted regression coefficients are shown in Table 5: for the full pre-intervention data, we report the mean, minimum, maximum and standard deviation of each coefficient over the 10 cross-validation folds, and for the other three sets of data we report the mean and standard deviation. The coefficient for the number of inhibitory behaviours is negative in all cases, so that the data does not support the EBIU hypothesis from Section 3: not every behaviour is associated with an increase in wellbeing.

The individual behaviours model brings another jump in predictive power; Table 6 gives the fitted coefficients. When the model is fitted to the full set of pre-intervention data, 13 behaviours have negative mean coefficients, which again does not support EBIU. The inhibitory behaviours appear disproportionately among those with the most negative coefficients. The pattern of results is similar for the other three sets of data.

Adding the number of opposite pairs as an extra predictor did not materially improve model performance; thus a simple OAS

[2]When modelling changes in wellbeing measures from pre- to post-intervention, the individual behaviour predictors are no longer Boolean, but can take three values: +1 if the behaviour was not reported pre-intervention but was reported post-intervention; -1 if the behaviour was reported pre-intervention but not post-intervention, and 0 if reporting of the behaviour was unchanged.

Table 6: Regression coefficients for the individual behaviours model fitted to the pre-intervention data for all participants, summarised across 10 cross-validation folds, with predictors ordered by coefficient mean on the full pre-intervention data. The "Group" column shows whether each behaviour is in the (F)acilitatory or (I)nhibitory group identified in Section 4.

Predictor (behaviour)	Group	Regression coefficients									
		Pre-intervention (all users)				Pre-int. (users with post- data)		Post-int.		Increase from pre- to post-	
		Mean	Min	Max	SD	Mean	SD	Mean	SD	Mean	SD
(Intercept)		439.8	437.9	442.1	1.15	459.1	2.22	489.5	2.37	40.7	0.58
Calm/relaxed	F	59.6	58.1	61.1	1.00	54.2	1.66	32.7	1.67	24.4	1.67
Energetic/driven	F	39.9	39.1	40.8	0.55	40.6	1.77	46.0	1.50	20.6	2.27
Definite	F	32.8	30.7	34.4	0.97	26.4	1.58	8.1	1.47	10.1	1.52
Flexible	F	24.4	23.0	25.8	0.80	33.5	1.92	30.3	1.71	19.3	1.01
Lively	F	20.5	19.3	22.0	0.92	18.1	1.41	11.5	2.06	20.7	1.06
Extroverted	F	17.7	16.9	19.2	0.65	16.3	1.74	17.3	1.45	2.6	1.66
Proactive	F	16.6	15.2	18.1	0.84	7.8	1.77	5.6	1.97	8.3	1.42
Systematic	F	15.8	15.0	17.3	0.60	18.9	2.05	14.5	1.33	2.2	1.48
Group-centred	F	14.5	12.8	16.9	1.15	9.6	1.51	7.4	1.67	1.4	1.30
Spontaneous	F	11.3	9.7	12.1	0.73	6.0	1.52	13.1	1.20	7.0	1.22
Behave as you wish	F	9.4	8.4	10.4	0.79	18.3	1.74	4.4	1.42	7.9	1.24
Conventional	I	6.1	4.8	7.8	0.95	11.3	2.20	10.3	2.06	2.3	2.62
Trusting	F	6.0	5.0	7.3	0.70	−1.1	2.06	16.1	1.97	0.6	1.52
Predictable	I	5.5	3.2	7.3	1.25	4.1	3.19	8.4	2.52	−4.4	2.02
Gentle	F	5.0	3.9	6.5	0.94	1.4	1.12	4.4	1.92	9.1	1.45
Open-minded	F	3.8	2.9	4.6	0.58	6.9	1.72	15.8	1.24	10.5	1.25
Risk-taker	F	1.0	−0.1	2.1	0.72	1.6	2.77	6.4	2.19	5.3	1.80
Firm	F	−0.2	−2.4	1.4	1.14	−4.3	1.94	9.7	1.77	9.1	1.41
Assertive	F	−0.5	−1.6	0.5	0.62	1.8	2.08	7.8	1.98	1.2	1.81
Single-minded	F	−1.0	−1.8	1.5	0.95	−2.0	2.29	0.1	2.14	−0.1	1.73
Individually-centred	F	−2.6	−3.6	−0.9	0.98	−9.2	1.83	−10.5	2.43	−4.7	1.97
Play it safe	I	−7.9	−9.2	−5.4	1.17	−12.9	1.77	−7.3	1.44	−17.4	2.00
Unconventional	F	−9.3	−11.2	−7.8	1.09	−6.1	1.56	−7.1	3.40	8.3	2.55
Unassertive	I	−18.3	−19.7	−17.2	0.73	−16.6	1.47	−18.4	2.39	−12.0	2.15
Behave as others want	I	−22.7	−23.7	−21.9	0.62	−17.3	1.96	−26.7	2.55	−12.2	0.66
Reactive	I	−23.5	−24.6	−22.2	0.89	−19.5	1.73	−23.4	1.96	−7.8	1.13
Introverted	I	−32.1	−33.2	−31.1	0.73	−44.4	1.56	−42.4	1.88	−22.9	2.55
Not lively / laid back	I	−37.3	−39.4	−34.4	1.49	−40.1	2.21	−30.1	2.49	−21.5	1.85
Wary of others	I	−43.0	−44.7	−41.6	1.11	−36.1	2.23	−28.4	2.24	−7.8	2.10
Unpredictable	F	−44.4	−46.1	−43.2	0.88	−40.9	4.52	−33.7	4.36	−5.4	3.30

hypothesis is not supported. Experiments with other opposites-based predictors also argue against OAS; we could find no evidence that selecting opposite behaviours confers any special benefit to an individual's wellbeing.

We have performed the same analyses using scores from an anxiety and depression diagnostic in place of wellbeing scores. In all cases the pattern of results was very similar.

We have used these findings to optimise the DSD "Do" selection algorithm. Up to 12 of the "Dos" selected for each person are expanders (described in Section 1) with a mean of 3.4 and a median of 4 expanders. Historically each person's expanders have targetted behaviours randomly chosen from those the person did not select on their pre-intervention behaviour rater.

We have now altered the algorithm so that behaviours which showed a consistently negative relationship with wellbeing in Table 6, and in similar models predicting anxiety and depression scores, will not be chosen unless the user has already selected all the other behaviours on their pre-intervention behaviour rater. We have also changed the selection probability for each target behaviour based on feedback elicited from a subset of users about the expanders; the selection probability is no longer uniform. Once sufficient programmes have been delivered using the new expander

Table 7: Grouping the data using quartiles of the pre-intervention wellbeing score.

Pre-intervention wellbeing score range	Number of participants	Response rate
$0 \leq$ score ≤ 392	3905	15.9%
$392 <$ score ≤ 496	3879	17.1%
$496 <$ score ≤ 594	3883	20.8%
$594 <$ score ≤ 800	3883	24.2%

Table 8: Grouping the data using quartiles of the pre-intervention behavioural flexibility score.

Pre-intervention flexibility score range	Number of participants	Response rate
$0 \leq$ score ≤ 11.7	4868	16.9%
$11.7 <$ score ≤ 16.7	3697	18.5%
$16.7 <$ score ≤ 23.4	3392	19.9%
$23.4 <$ score ≤ 100	3593	23.7%

selection algorithm, we will conduct a comparison between the new algorithm and the earlier one.

6 INVESTIGATING RESPONSE RATES AND APPLYING PROPENSITY SCORES

We noted in Section 2 that in terms of pre-intervention wellbeing and behavioural flexibility scores, the subset of participants completing the post-intervention questionnaire is not representative of the full population of those enrolling: those completing the post-intervention questionnaire have somewhat higher initial wellbeing and flexibility scores (see Table 1). This raises the prospect of nonresponse bias.

The main conclusions we have presented so far are not endangered by this because they are established on the full set of pre-intervention data. For example, the correlation network in Section 4 was derived from the full pre-intervention data, and the main findings from Section 5 are supported both by the full pre-intervention data and by the other sets of data[3].

Nevertheless, from an intervention design and evaluation perspective it is worthwhile to investigate the factors that influence the likelihood of post-intervention questionnaire completion, and also how these same factors relate to the average benefit experienced by participants. We do so in this section, using the method of *propensity scores* (explained for example in [1, 9, 13]).

Nonresponse to the post-intervention questionnaire can happen in two distinct ways:

[3]The relative performance of the various models is very consistent across the four sets of data, as shown in Table 4. The regression coefficients given in Table 5 for the facilitatory/inhibitory model have very similar values for the full pre-intervention data, the pre-intervention data for those users with post- data available, and for the post-intervention data. Table 6 shows a similar ranking of behaviours across all four sets of data, with the inhibitory behaviours being always disproportionately represented among those with the most negative coefficients.

Table 9: Grouping the data using quartiles of the number of facilitatory behaviours selected pre-intervention.

Pre-intervention number of facilitatory behaviours	Number of participants	Response rate
0 to 4	4234	16.7%
5 to 6	3585	18.7%
7 to 9	4297	19.5%
9 to 21	3434	23.8%

Table 10: Grouping the data by the number of inhibitory behaviours selected pre-intervention.

Pre-intervention number of inhibitory behaviours	Number of participants	Response rate
0	3322	20.1%
1	3504	19.3%
2	2911	18.7%
3	2157	17.9%
4	1555	19.7%
5 to 9	2101	21.5%

(1) A participant can complete the actions recommended in the "Dos" they are sent, but then not complete the post-intervention questionnaire.
(2) After enrolling a participant can decide, for whatever reason, not to follow the intervention i.e. not to complete their recommended "Dos".

Because "Dos" are small actions the participant completes on their own, we cannot know whether they really carried them out, and thus we cannot distinguish between the two kinds of nonresponse. However, we emphasize that intervention designers would like to reduce nonresponse regardless of which source it comes from.

Table 7 groups the data using the quartiles of the pre-intervention wellbeing score, showing the response rate in each quartile. Table 8 shows the corresponding breakdown using quartiles of the pre-intervention flexibility score. We see that people with better pre-intervention wellbeing are more likely to complete the post-intervention questionnaire, as are people with a higher pre-intervention flexibility score. Table 9 shows that the number of facilitatory behaviours selected pre-intervention is similarly predictive of response rate. By contrast, the number of inhibitory behaviours selected pre-intervention does not exhibit a monotonic relationship with response rate, as shown in Table 10.

To apply the method of propensity scores, we fit a logistic regression model of post-intervention questionnaire response using as predictors four pre-intervention variables — wellbeing score, flexibility score, number of facilitatory behaviours selected and number of inhibitory behaviours selected — and all pairwise interactions between them. The propensity score for each individual is the prediction of this model, between 0 and 1. Informally, people with a low propensity score are the type of people who are relatively unlikely to complete the post-intervention questionnare (based on their pre-intervention questionnaire answers); people with a high propensity score are the type of people who are relatively likely to complete it.

Following common practice, we now stratify our data into five groups, using the quintiles of the propensity scores, and examine what is happening in each of the five groups. Table 11 shows the response rate and the changes in wellbeing in each group. We emphasise that *within each of the five groups*, the subset of users completing the post-intervention questionnaire is (almost) representative of the full set of users in the group. For example, consider the 4th and 5th columns of Table 11, which show respectively the mean pre-intervention wellbeing score of each group, and the mean pre-intervention score of just those users who completed the post-intervention questionnaire. These values are well matched for each group, with the largest discrepancy being 4.6 (as shown in the "Absolute difference" column). By contrast, when we do not use stratification but work with the whole dataset, there is a much larger mismatch of 23.1 (i.e. the difference between 486.9 and 510.0 in Table 1).

Although we present this assessment of representativeness for the mean wellbeing score only, we have verified that the match within each of the five groups is reasonably good also for the mean pre-intervention flexibility score, and the mean numbers of facilitarory and inhibitory behaviours selected pre-intervention. Furthermore the standard deviations of these four variables (in addition to their means) are also well matched. The fact that propensity scores allow us to achieve this kind of balance across multiple variables simultaneously is a key reason for using them [12]; the propensity score gives a convenient single number that "contains information about all the measured covariates summarized into a single variable that researchers can use to stratify patients" [13].

The main conclusion we draw from the results of Table 11 is that the kinds of people who are the least likely to tell us they have completed the intervention by filling in the post-intervention questionnaire, are also the kind of people who will experience the largest increase in wellbeing if they do complete it. At the other end of the spectrum, the kinds of people who are most likely to tell us they have completed the intervention are the kind of people who will experience the smallest increase in wellbeing if they do complete it. This may be partly due to the fixed range of the wellbeing questions: someone who provided a pre-intervention answer close to the upper bound of 100 on one of the questions cannot report much of an improvement post-intervention. In any case, an increase in mean wellbeing in evident within all the five groups.

7 CONCLUSIONS

By applying data mining techniques to a large dataset of answers to a behaviour and wellbeing questionnaire, collected from participants in digital behaviour change interventions, we developed a model of how behaviours are linked to wellbeing that fits the data much better than the existing behavioural flexibility formula.

In particular, by constructing correlation networks we found that the 30 behaviours included in the DSD system break down into two meaingful groups: the facilitatory behaviours and the inhibitory behaviours. Using regression modelling, we found that while the majority of the 30 behaviours included in the DSD model are associated with better wellbeing, a number of them are associated with poorer wellbeing. These negatively associated behaviours contain most of the inhibitory group, and regression models using the facilitatory/inhibitory distinction explain wellbeing much better than the existing formula.

Our improved model thus suggests that rather than increasing the *number* of behaviours people have in their repertoire, the behavioural part of DSD interventions works by helping people *switch* their behaviours from inhibitory ones, which reduce their wellbeing to facilitatory ones which increase their wellbeing. The findings we have presented are among those we have used to optimise DSD behaviour change interventions, ensuring that interventions concentrate on helping people to develop the behaviours that are likely to benefit them most.

By stratifying our dataset using propensity scores, we found that the kind of people who are the least likely to tell us they have completed the intervention, by answering the post-intervention questionnaire, are also the kind of people who will experience the greatest increase in wellbeing from a completed programme.

Our results about behaviours and wellbeing will also be of broader interest, given the relatively large size of our dataset. In a widely cited meta-analysis [14] of research on how wellbeing is affected by personality traits such as introversion and extroversion, the median number of people included in each such study was 179 and the mean was 354 (for the 357 studies analysed). By contrast we analysed the data for 15,550 people.

ACKNOWLEDGMENTS

Charlton, Kingston and Petridis were supported by Innovate UK via Knowledge Transfer Partnership 10152 (a Knowledge Transfer Partnership between the University of Brighton and Do Something Different Ltd). Fletcher was supported by EU Horizon 2020 grant 643735 (http://www.do-change.eu/about/) "Do Cardiac Health: Advanced New Generation Ecosystem". We thank the anonymous referees for their helpful suggestions.

Competing interests. Do Something Different Ltd offers some of its behaviour change programmes on a commercial basis. No other competing interests exist.

REFERENCES

[1] Peter C. Austin. 2011. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research* 46, 3 (2011), 399–424.

[2] Giulio Costantini, Sacha Epskamp, Denny Borsboom, Marco Perugini, René Mõttus, Lourens J. Waldorp, and Angélique O.J. Cramer. 2015. State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality* 54 (2015), 13 – 29. R Special Issue.

[3] Angélique O. J. Cramer, Sophie van der Sluis, Arjen Noordhof, Marieke Wichers, Nicole Geschwind, Steven H. Aggen, Kenneth S. Kendler, and Denny Borsboom. 2012. Dimensions of Normal Personality as Networks in Search of Equilibrium: You Can't Like Parties if You Don't Like People. *European Journal of Personality* 26, 4 (2012), 414–431.

[4] Ben (C) Fletcher, Jill Hanson, Karen J. Pine, and Nadine Page. 2011. FIT - Do Something Different: A new psychological intervention tool for facilitating weight loss. *Swiss Journal of Psychology* 70, 1 (2011), 25–34.

[5] Ben (C) Fletcher and Karen J. Pine. 2012. *Flex: Do Something Different.* University of Hertfordshire Press.

[6] Ben (C) Fletcher, Karen J. Pine, and Danny Penman. 2005. *The No Diet Diet: Do Something Different.* Orion Publishing Group, London.

Table 11: Changes in wellbeing, stratified by propensity score.

Propensity score range	Number of users	Response rate	Pre-intervention wellbeing (all users)	Pre-intervention wellbeing (users with post-data available)	Absolute difference	Post-intervention wellbeing	Increase in wellbeing from pre- to post-
$0.081 \leq p \leq 0.160$	3110	14.9%	302.9	304.3	1.4	441.9	137.6
$0.160 < p \leq 0.182$	3111	15.7%	432.2	434.0	1.8	514.2	80.1
$0.182 < p \leq 0.202$	3109	19.5%	504.3	508.9	4.6	557.9	49.0
$0.202 < p \leq 0.227$	3110	21.8%	564.6	568.9	4.3	595.0	26.1
$0.227 < p \leq 0.553$	3110	25.6%	630.5	627.3	3.3	633.1	5.9

[7] Ben (C) Fletcher and Bob Stead. 2000. *(Inner) FITness & The FIT Corporation.* International Thomson Press, London.

[8] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (second ed.). Springer.

[9] Marshall M. Joffe and Paul R. Rosenbaum. 1999. Invited Commentary: Propensity Scores. *American Journal of Epidemiology* 150, 4 (1999), 327–333.

[10] Karen J. Pine and Ben (C) Fletcher. 2011. *Love Not Smoking: Do Something Different.* Hay House UK.

[11] Karen J. Pine and Ben (C) Fletcher. 2016. *Changing people's habits is associated with reductions in stress, anxiety and depression levels.* Technical Report. Do Something Different Ltd. http://dsd.me/healthcare/the-evidence/

[12] Paul R. Rosenbaum and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 1 (1983), 41–55.

[13] Kristin L. Sainani. 2012. Propensity Scores: Uses and Limitations. *PM&R* 4, 9 (2012), 693–697.

[14] Piers Steel, Joseph Schmidt, and Jonas Shultz. 2008. Refining the relationship between personality and subjective well-being. *Psychological Bulletin* 134, 1 (2008), 138–161.

[15] Ruth Tennant, Louise Hiller, Ruth Fishwick, Stephen Platt, Stephen Joseph, Scott Weich, Jane Parkinson, Jenny Secker, and Sarah Stewart-Brown. 2007. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes* 5, 1 (2007), 63.