The BIG measure - Results

2023-10-05

# Demographics

We ran two separate cohorts through our study during the fall semester of 2022 and the winter semester of 2023.

For Cohort 1 we had 87 subjects complete the Intention-Behavior gap measure at onboarding and 77 at offboarding.

For Cohort 2 we had 93 subjects complete the Intention-Behavior gap measure at onboarding and 85 again at offboarding.

In total we had 180 subjects complete the Intention-Behavior gap measure at onboarding (36 males, 124 females, 3 other) with a mean age of 18.79 (SD = 1.34).

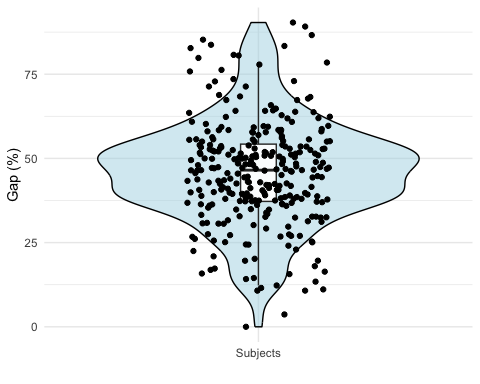
The majority of particpants were Asian (55.56%), followed by Other (17.22%), Black or African American (11.11%), White (6.11%) and Indigenous or Native (0.56%).

# Gap Characteristics

A simple domain-general gap measure was calculated by averaging the magnitude of the gap for all goal domains each subject selected. Later, we created a “weighted” version of the gap, leveraging the fact that we collect additional data from our subjects about each goal domain (e.g. domain importance, required effort, time, etc.). This is explained in detail in section 4.4 (Informing the Intention-Behavior Gap).

## Gap distribution in population

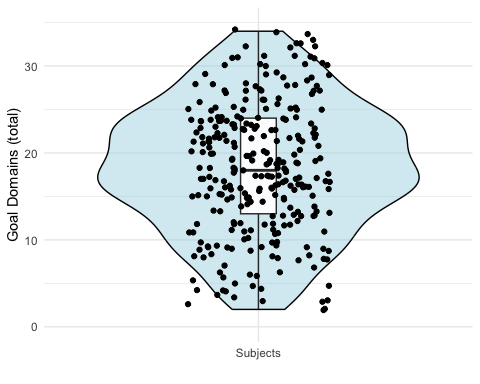
Ignoring this additional information for the moment, we can look at the “unweighted” measure of our gap and observe that the distribution of averaged gap magnitudes looks close to normal (Figure @ref(fig:gap-distribution)), with a mean gap across subjects of 45.9% (SD = 15.8). We can also see that a single very productive subject indicated that they had a gap of zero.



Distribution of Intention-Behavior gap magnitudes in sample.

## Goal Domain Count

Given that one unique aspect of our measure is that subjects can effectively choose the number of items (i.e. goal domains) based on self-relevance (up to a maximum of 34), we necessarily end up with variation in the total number of items/goals for each subject. Across subjects we can again see that this number approximates a normal distribution (Figure @ref(fig:goal-domains-count)), with a mean of 18.3 goal domains selected (SD = 7.5).



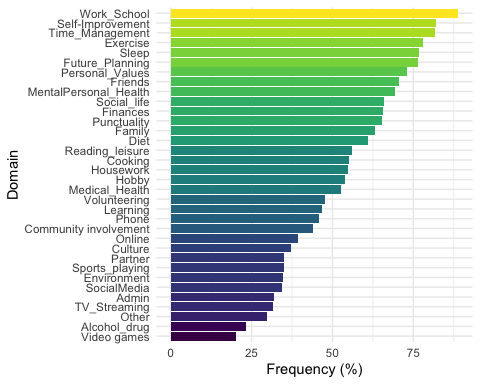
Number of goal domains for each subject.

## Goal Domain Characteristics

### Frequency

To provide an overview of some of the characteristics of our different goal domains we can look at domain selection frequency, importance, and motivation source (to what degree a goals is internally vs. externally motivated) for all of the domains.

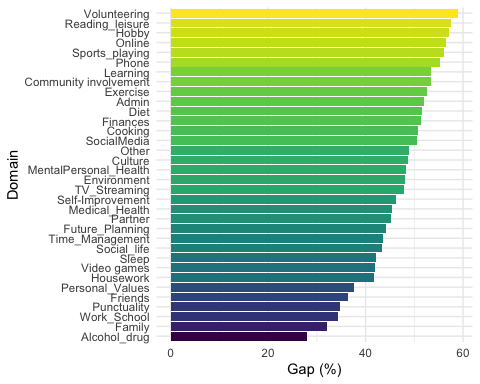
The frequency with which domains were selected as “goal” domains varied greatly across goal categories, with work/school being chosen most commonly, 89% of the time, and video games least, just 20% (detail in Figure @ref(fig:goal-domains-frequency)).



Goal domain frequency of selection.

### Magnitude

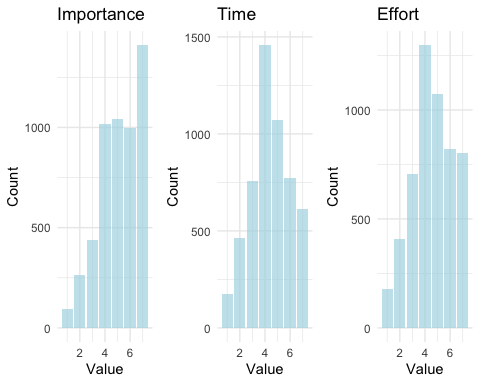
Given that we had so many different goal domains (34) we were curious how the magnitude of the gap might differ by domain. As shown in Figure @ref(fig:goal-domains-gap) we can see that there is good deal of heterogeneity in the gaps (where 100% would indicate not accomplishing a goal at all, and 0% would be complete accomplishment). The largest gap (volunteering, 59%) was in fact just over twice the magnitude of the smallest gap (alcohol and drug use, 28%).



Gap magnitude by goal domain.

### Additional Domain Information

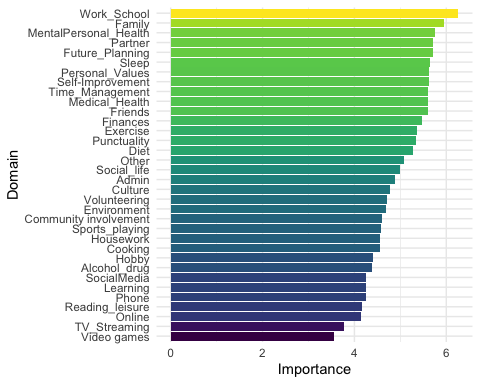
As mentioned, we collected data on a number of dimensions that we thought could provide important additional information about each goal domain, including importance, time (i.e. how much time is required to accomplish the goal takes), and effort. We plot those distributions below (Figure @ref(fig:intention-importance-distribution)). These specific dimensions were later used to create a “weighted” Intention-Behavior gap (See the section [Informing the Intention-Behavior Gap](#informing-the-intention-behavior-gap)).



Distributions of importance, time, and effort evaluations for subject goal domains as selected using a 7-point Likert scale.

#### Goal Importance

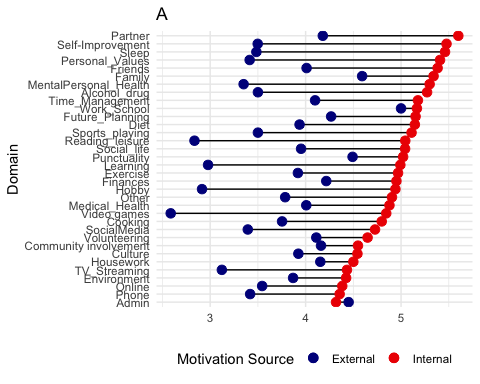
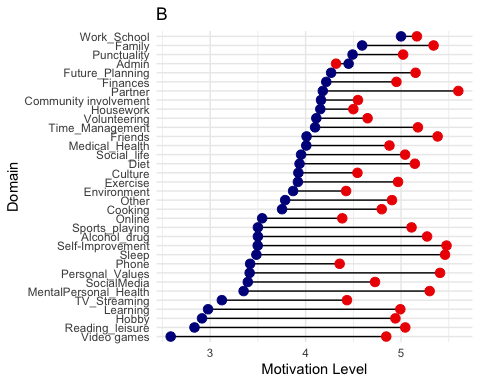
When considering importance scores we can see that domain means only fall in the upper half of the possible range (Figure @ref(fig:goal-domains-importance)), with the domain receiving the lowest score, video games, still received an average score slightly higher than the midpoint of the range (M = 3.55, SD = 2.06). This compression of the range we believe is likely due to the fact that subjects self-selected those domains for which they felt like they currently held a goal. The effect of this is that any domain with an importance of or close to zero would be eliminated. The domain rated most important (M = 6.26, SD = 0.98) was also the domain where subjects most commonly had a goal: Work/School.



Average goal importance by domain, measured using a 7-point Likert scale.

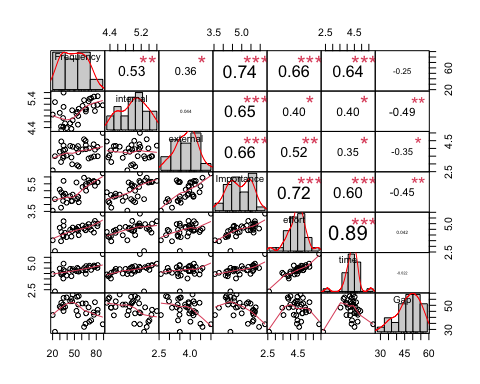
#### Locus of Motivation

For each goal domain a subject selected as self-relevant we then asked them to indicate to what degree their goal in the chosen domain was internally and externally motivated. We can look at the difference in these scores (internal motivation - external motivation) as a measure of how much more internally motivated each domain is. We find that all domains, in fact, are more internally motivated than externally motivated, save for administrative work. We also note the interesting finding that while the Learning goal is highly internally motivated, the work/school goal is the second least internally motivated domain. For our sample we know that all subjects are students, but we are also aware that some subjects also work an additional job. This means that even though this category is mostly representing school, there is some contamination by people who are also working an outside job.

### Relationships between gap, Domain frequency and Domain dimensions

We can also look at the relationships between the different goal domain attributes and Intention-Behavior gap magnitude.

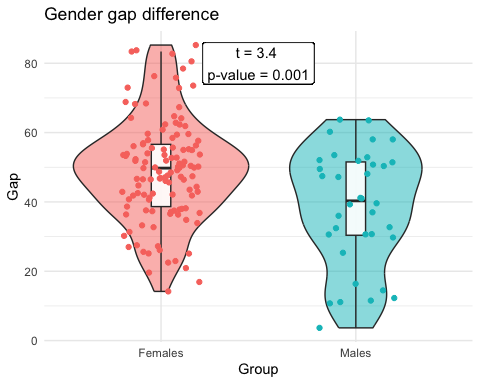


Correlations between goal domain attributes. Note: \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

We observe that the variables with the strongest associations (all negative) with gap magnitude are internal motivation (r(32) = -0.49, p = 0.003), goal importance (r(32) = -0.45, p = 0.008), and external motivation (r(32) = -0.35, p = 0.04). It is also interesting to note in passing that while more highly internally motivated goals tend to take more time (r(32) = 0.4, p = 0.02) than externally motivated goals (r(32) = 0.35, p = 0.04), they are less correlated with effort (r(32) = 0.4, p = 0.018) than externally motivated goal levels (r(32) = 0.52, p = 0.002).

## Gaps by Gender

We found a significant difference in the magnitude of the reported Intention-Behavior gap for females (M = 49.06%, SD = 15.01) and males (M = 38.5%, SD = 16.79). A Welch two-samples t-test showed that the difference was statistically significant, t(52.76) = 3.4, p = 0.001).



# Reliability

Reliability in the context of a psychological measure refers to the consistency, stability, and repeatability of the scores produced by the measure. When an instrument is reliable, it yields consistent results under consistent conditions. This doesn’t necessarily mean the instrument is measuring what it’s supposed to measure (that’s validity), but that it measures consistently (Nunnally, J. C., 1976).

## Cronbachs Alpha

Cronbach’s Alpha () is a statistic commonly used to measure the internal consistency of a scale or test, often in the fields of psychology, education, and related disciplines. In essence, it assesses how well a set of items (questions, tasks, etc.) measures a single construct. The value of ranges between 0 and 1, with higher values indicating greater internal consistency (Tavakol, M. & Dennick, R., 2011).

Mathematically, Cronbach’s Alpha is given by:

Where represents the number of items, is the variance of item , and is the variance of the observed total scores.

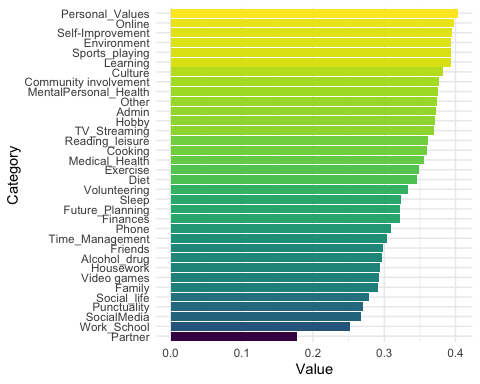
We calculated Cronbach’s alpha using a dataframe that included many cells with NA values, given that participants generally did not have goals for all domains. In total missing values were present in 46.3% of the dataframe’s cells. Even with a sparse matrix the psych library (v2.3.6, Revelle) can still calculate a Cronbach’s alpha metric by removing missing values and calculating pairwise correlations. Using this library internal consistency as measured by Cronbach’s (standardized) was 0.95.

While this supports the idea that we have an internally consistent uni-dimensional measure, we should also be aware that the value is inflated due to the high number of items (in our case 34) in the measure (Taber, K. S., 2018). One interpretation of a very high alpha value is that the items might be redundant. In our case, given the nature of our items, this is not a major concern as we have specifically chosen distinct domains. For example, exercise and diet may be closely related for many people, but they are not the same activity.

## Inter-Item Correlation

Inter-item correlation refers to the pairwise correlations between items on a measure as a means of evaluating the consistency of the measure. If items on a scale are supposed to tap into the same underlying construct, they should correlate positively with each other (Nunnally, J. C., & Bernstein, I. H., 1994).

This metric is calculated by computing a correlation matrix for all items and then averaging the correlations of the matrix (excluding the diagonal which will have a perfect correlation of 1 with itself). In other words, given a correlation matrix where represents the correlation between item and item , and is the number of items:



Average pairwise correlation of each goal domain with all other goal domains.

The mean inter-item correlation for our measure is 0.34, which is within the ideal range of .20 to .40, indicating that items are neither so different from one another that they are unlikely to be touching on the same measurement domain, nor so homogeneous that they are redundant (Piedmont, 2014).

The highest single bivariate correlation is between ib\_domain\_success\_Video games and ib\_domain\_success\_Alcohol\_drug of r(18) = 0.78, p < 0.001. This underscores why for this measure a strong correlation does not suggest that items are equivalent.

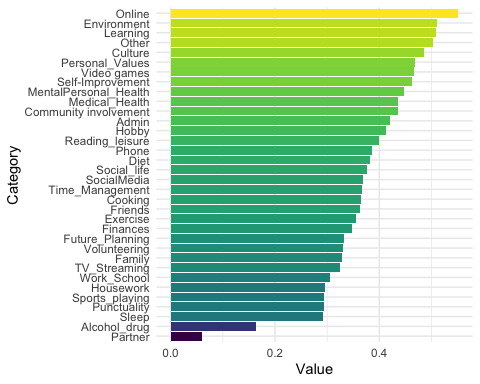
There are also two negative correlations between items - both involving video gaming intentions. The first is with intentions toward the subject’s relationship with a partner (), and the second with community involvement goals (), however neither was significant.

## Item-Total Correlation

Item-total correlation, often referred to as the corrected item-total correlation, represents the correlation between a particular item and the sum of all the other items in a scale or test. The “corrected” aspect means that the total doesn’t include the item itself. It is commonly used in psychometric analyses to gauge how well an item aligns with the overall scale or test.

We calculated corrected item-total correlations by testing how the score of an individual item correlates with the total score of all items, excluding the score of item . It provides a measure of how much the item relates to the rest of the test when the influence of the item itself is removed from the total score. The full formula is:

Where is the mean score of item and is the mean of the corrected scores ().



Corrected item-total correlation of each domain with sum all other domains, excluding itself.

The average corrected item-total correlation was 0.38. This represents 13 items that had item-total correlations ≥.4, which indicates excellent discrimination, 15 items that were between .3 and .4, which indicates good discrimination, four items that were between .2 and .3 which indicates marginal discrimination (though all were above .29) and two items ≤0.19 which means poor discrimination (Qin, 2006, Streiner & Norman, 2008).

## Test-Retest Reliability

Test-retest reliability refers to the extent to which scores on a particular measure are stable over a specified period. In other words, if the same participants complete the same measure on two different occasions (with no intervention or change occurring between the two test times), their scores should be similar if the measure is reliable. This is particularly true for constructs that are expected to be stable over time, such as intelligence or personality. However, when measuring constructs that might be expected to change over fairly short time periods (e.g. mood), the test-retest method can be less appropriate as a measure of reliability (Nunnally, J. C., & Bernstein, I. H., 1994, Streiner, D. L., & Norman, G. R., 2008).

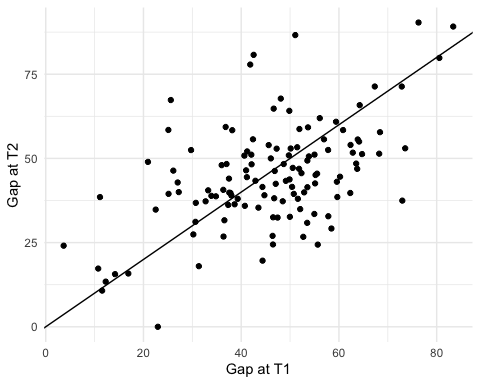
We use the Intraclass Correlation Coefficient (ICC) to quantify test-retest reliability. This has been found to be superior than simply calculating the Pearson correlation between scores at two different times as it assesses not just the correlation, but also the agreement between measurements. There are a number of versions of the ICC, we use ICC(3,1) which is used for single measures and based on a two-way mixed-effects model. It is calculated by:

Where is the between-subjects variance, and is the within-subjects variance (McGraw, K. O., & Wong, S. P., 1996).

### Test-Retest Evaluation

We had 125 subjects complete the measure twice - once near the beginning of semester (t1), and a second time (t2) about 3 months later at the end of the semester (weeks = 13.7, SD = 4.6).

The ICC for the two measurement timepoints was found to be ICC(3,1) = 0.53, 95% CI [0.39, 0.64], indicating moderate reliability (Koo and Li 2016). For reference the Short Grit Scale had a 1 year test-retest stability of and conscientiousness scores based on the NEO Five-Factor Inventory (Costa & McCrae, 1992) correlated at over 4 years in a test by Robins, Fraley, Roberts and Trzesniewski (2001). Notably, within our own data we found that most comparison trait measures had a higher ICC value than our gap measure at t1 and t2 (see Table @ref(tab:test-retest-other-measures)).



Each dot represents one subject’s Intention-Behavior gap score at timepoint one (t1) and timepoint two (t2).

We believe this relatively low correlation for the gap measure could be partially due to influence on responses from current state levels of the gap at the time of completing the measure, which are expected to fluctuate. Additionally, we are studying a population (almost all first year undergraduate students) that we would expect to have higher state-level variance in their Intention-Behavior gap due to the unique moment in their lives starting a university education represents. As with any major life change, novel situations are ripe ground for Intention-Behavior gaps as we are attempting to set goals we have not previously attempted. One might expect someone in their 30s or 40s with a regular job to have a more stable test-retest value for this measure.

Comparison of the test-retest reliability of the Intention-Behavior gap measure with other moderator and outcome variables.

Measure

ICC

LowerBound

UpperBound

Intention-Behavior Gap

0.53

0.39

0.64

Self-Control

0.74

0.71

0.76

Conscientiousness

0.74

0.71

0.76

Grit

0.66

0.63

0.70

DASS

0.65

0.61

0.69

Subjective Happiness

0.68

0.65

0.71

Quality of Life

0.61

0.57

0.64

Flourishing

0.66

0.62

0.69

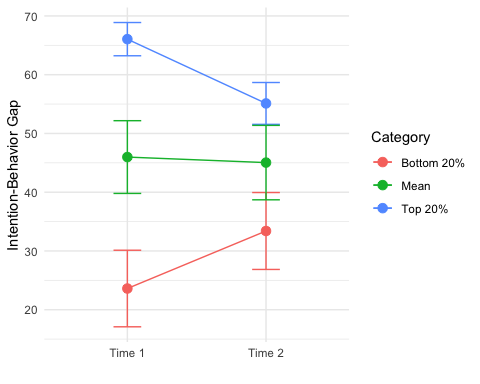
### Test-Retest Variability

We decided to test the potential influence of both regression to the mean and current state-level gaps on the trait level measure. These two approaches are related in that we would expect the mechanism behind a regression to the mean to be largely driven by the fact that individuals happened to be in a particularly high or low-gap state at the time of completing the trait measure.

#### Regression to the mean

We first tested if regression to the mean did, in fact, seem to be occurring in our sample. Regression to the mean describes the general phenomenon that within a given subject particularly high or low measurements tend to be followed by measurements that are closer to the population mean (Barnett, Pols & Dobson, 2005). The explanation for this phenomenon is that sampling variability itself tends to be normally distributed so that within subject if you take a measurement that is at the far end of their distribution, due to noise, the next time you sample that same person it is unlikely that you will get such an extreme value. To see if this was the case with our own data we took the top an bottom quintiles of our Intention-Behavior gap measure scores at t1 and looked at how the scores for those same subjects changed when moving to t2.

We did find support for our intuition that regression to the mean was occurring (see Figure @ref(fig:test-retest-regression-to-mean-plot)). Taking the highest 20% of gaps from t1 the average was a 66.05% gap (SD = 6.93). At t2 this number had decreased on average to 55.11% (SD = 15.96). The same pattern, but in the opposite direction, was found when looking at the lowest 20% of gaps, where the t1 average was 23.61% (SD = 8.72) and the t2 average 33.4 (SD = 16.03). The difference in means was significant in both cases, based on a one-tailed Welch two sample t-test (p = 0.002 in the first case and p = 0.005).



Change in average Intention-Behavior gap scores for the top and bottom quintiles of subjects as measured at t1.

#### Recency bias

Given that there does seem to be regression to the mean in our samples, we can question what might be causing this measurement error. One possible source, as pointed out before, is the biasing of responses that are intended to capture trait-level gap magnitude by a subject’s current state-gap level gap. Given that for most subjects we have daily measurements of their state-level Intention-Behavior gaps we can look to see if there seems to be an influence from temporally adjacent state-level gap reports and the trait-level measure.

To test our hypothesis we took the completion date of each subject’s trait-level measure at t1 and t2 and compared it to an average of state-level gap reports for the week leading up to the measurement (including the day of the measurement). Note that in some cases we did not have data for all seven days, and so simply averaged across the days that were accounted for in that week-long window. Our expectation was that we would find a higher correlation between the trait-level measure and the temporally adjacent state-level seven-day average, than between the trait-level measure and a more temporally distant state-level seven-day average (i.e., ).

Note that when there was missing data (i.e. not all measurements for the seven day period were present) we conducted the calculation on the reduced number of days rather than keep going back in time, as the state-level correlation would be expected to decay. We also did not create state-level scores for people who completed their off-boarding trait-level measures more than 3 days after the completion of data collection. Again, we wanted to avoid having an extended gap between the trait-level measure and the state-level data.

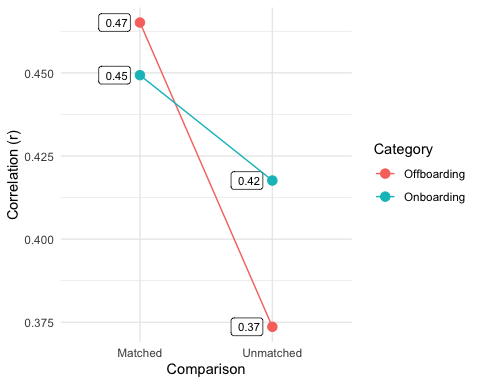
We used Fisher’s z-transformation to test whether the two correlations were significantly different, where:

And standard error () is calculated as follows:

Where and are the sample sizes for the two correlations. Fisher’s z-transformation is described as:

We then conducted a one-tailed test for significance:

We found, as anticipated, that the daily-gap measures that were proximate in time to the onboarding Intention-Behavior gap measure correlated slightly more strongly with the initial gap measures than with the offboarding Intention-Behavior gap measures, but the difference was not significant (). The daily-gap measures proximate in time to the offboarding measure were also more strongly correlated with the offboarding Intention-Behavior gap measure than with the onboarding measure, as predicted, but, again, those differences were not significant (). So, while the hypothesized effect is strictly present, the magnitude of that effect is not large enough to allow us to say that our hypothesis has been supported by the data (Figure @ref(fig:test-retest-state-effects-plot)).



Comparison of correlation values of temporally matched state and trait measurments, versus temporally distant state and trait measures.

## Measure vs. Daily Mean

Given that we collected 12 weeks of daily measures of the Intention-Behavior gap, another opportunity to assess the reliability of the Intention-Behavior gap measure would be to see how an average of the onboarding and offboarding Intention-Behavior gap trait measures correlates with the daily reported gaps over the course of the 12-week study. This test of reliability is motivated by the idea that you would expect many measurements of the state-level Intention-Behavior gap taken over an extended period of time to approximate the trait level measure. We tested to see if this was the case in our data, with the caveat that since we do not have a clear idea of the timescale of fluctuations in the state-level measure, it is not guaranteed that 12 weeks is a long enough time period to reliably approach each subjects true trait-level mean. To conduct our test we used an intraclass correlation statistic to test how similar the mean of the daily self-reports were with the Intention Behavior Gap measure value for each participant. As a reminder, the daily self-report measure was provided by participants each evening. The number of self reports varied, up to a max of 84 ().

In the table below (Table @ref(tab:measure-dailyGap-table)) we looked at both the onboarding and offboarding measurements separately as well as a combination of the two measurements (averaged together). We also looked at the daily gap as calculated based on a single measure (“Over the past 24 hours the level of my intention gap was:”, on a scale of 0-100%), as well as a composite measure which included the gaps on two specific goals the participant had set for themselves. We used a simple weighting scheme of 75%/25% for the overall measure and the two specific goal measures. In all cases we found that this composite daily gap measure was more highly correlated with the Intention-Behavior Gap measure, and also found that combining both the onboarding and offboarding measurements of the instrument provided the highest correlation.

ICC of Intention Behavior Gap measure and daily self-report of gap

Measure Source

Daily Gap

ICC

onboarding

single measure

0.47

onboarding

composite

0.51

offboarding

single measure

0.54

offboarding

composite

0.60

combined

single measure

0.56

combined

composite

0.62

# Validation

In order to assess how well our measure is accurately capturing a domain-general Intention-Behavior gap we employed both traditional and measure-specific methods. We will start with examining convergent validity, where we test associations between our measure and other measures that we believe should, theoretically, be associated with the Intention-Behavior gap. We then conduct a test to check that our measure predicts out of sample domains for a given subject better than the population mean. We then test whether we can gain increased explanatory power by weighting our Intention-Behavior gap using additional information we have for each domain (importance, time and effort). We identify models of best-fit for our outcome measures from a set of predictors including all moderators and our weighted Intention-Behavior gap measure. Finally, we confirm that the Intention-Behavior gap is indeed related to behavior, and not simply ambitious goal-setting.

## Content Validity

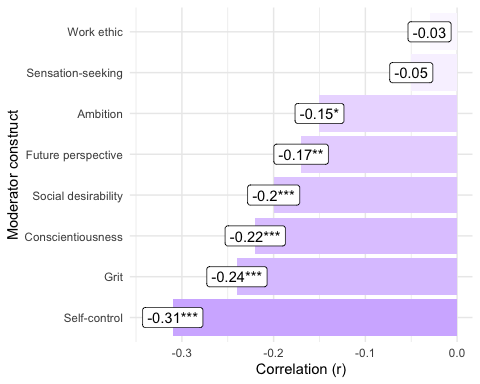
We attempted to establish content validity of our measure, the idea that our items captures the full range of the construct/characteristic, by first building off canonical instruments. Specifically we used the Canadian Time Use Survey (2015 - 2016, Cycle 29) developed by Statistics Canada and the American Time Use Survey (2011-2022) developed by the U.S. Bureau of Labor Statistics. We then used the Delphi method to refine and add to this list.

Beyond these steps, given the phenomenon of “unknown-unknowns” it is tricky to confirm that a measure is sampling from the entirety of the domain the concept in question covers. However, in our case we included an optional “other” goal category in the measure, which could be seen as an indicator of the degree to which our items were incomplete. A minority of subjects, 29.9%, indicated that they had a goal that did not fit within our listed categories. We did not require participants to specify what their “other” goal was, so while we may assume that their responses would have been idiosyncratic enough not to merit a single new category, we are unable to confirm this, which is a shortcoming of the study design. That said, for the majority of subjects the provided items appeared to capture the full range of their goal categories.

## Convergent Validity

### Moderators

Given that there are not established measures of the Intention-Behavior gap that we are aware of, we were not able to compare the performance of our instrument to others that are designed to measure the same construct. That said, we did select a number of measures that we hypothesized would moderate the translation of intentions to behavior, specifically self-control, conscientiousness, grit, sensation-seeking, future time perspective, ambition, social desirability bias, and work ethic.



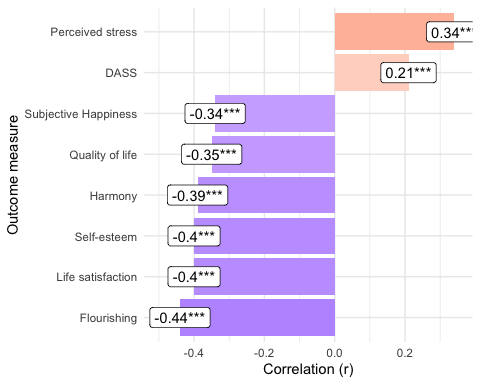
Correlations between the Intention-Behavior gap and constructs believed to moderate the relationship between intentions and behavior. Note: \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

We found that all hypothesized moderators, except for sensation seeking and work ethic, were significantly correlated with the Intention-Behavior gap (see Table above). Those measures that were significantly associated had small to moderate correlations. This is in line with our expectations as we do not see these constructs as equivalent to the gap, and we also believe there will be considerable effects of individual differences in terms of the strength of any particular moderator. This level of correlation also reduces the risk of issues arising due to multicollinearity when conducting model comparison.

### Outcomes: Well-being

In addition to moderators of the gap, we hypothesized that the Intention-Behavior gap would be negatively associated with well-being (i.e. a higher level of gap would correlate with lower levels of well-being). To test this hypothesis and provide further validation for our measure we looked at measures for flourishing, harmony, quality of life, satisfaction with life, subjective happiness, self-esteem, stress, and a composite depression/anxiety/stress scale. It is worth noting that given the estimate that intentional activity accounts for approximately 40% of the variance in happiness (Lyubomirsky, Sheldon, & Schkade, 2005), theoretically we would expect an upper bound of any given correlation with well-being to be .

We did find that, as hypothesized, all of our measures of well-being were significantly correlated () with the Intention-Behavior gap, most at a moderate level (see Figure @ref(fig:predictive-validity-plot)).



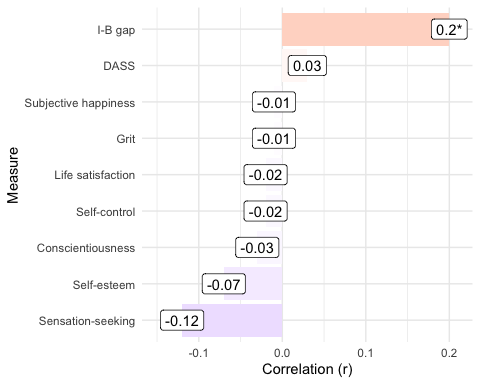
Correlations between the Intention-Behavior gap and well-being measures. DASS refers to the Depression, Anxiety and Stress Scale. Note: \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

### Outcomes: Empirical

In addition to non-observable psychological constructs (such as well-being), we looked at two empirical outcomes, body mass index (BMI, self-reported) and grades (provided by the University of Toronto).

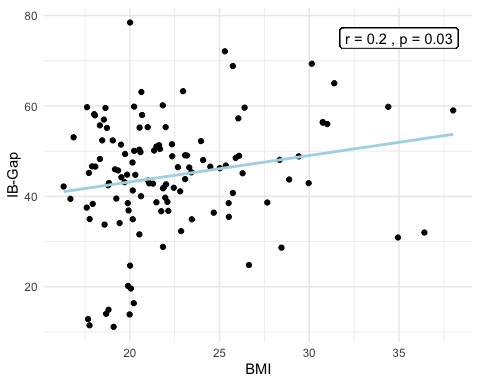
#### BMI

We looked at associations between BMI and a number of moderator and outcome variables along with the Intention-Behavior gap.



Correlations between various measures and BMI. Note: \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

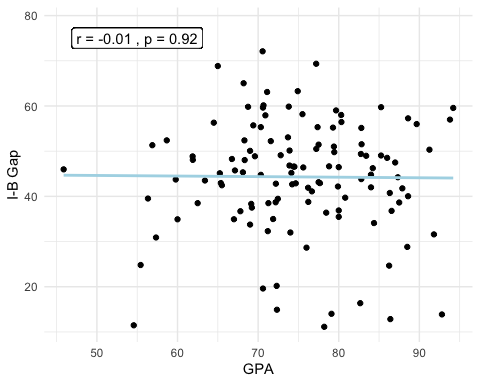
We found that only the Intention-Behavior gap was significantly associated with BMI. The association was small and positive (i.e. a larger gap correlates with a higher BMI). While most of the remaining measures tended to at least correlate in the anticipated direction (e.g. DASS: positive, self-control: negative), none of those correlations reached significance (). We can also view the scatterplot of individual subjects which illustrates this relationship more clearly (Figure @ref(fig:bmi-measure-scatterplot)).



Scatterplot of subjects BMI versus their Intention-Behavior gap measure. Note that both BMI and the gap measure are calculated from an average of the t1 and t2 values, for those subjects that responded at both timepoints.

#### Grades

We collected end of semester grade data from all participants and then tested for relationships between the same set of variables as chosen for BMI. In this case if we plot our subjects as before (Figure @ref(fig:grades-gap-scatterplot)) we can see that there is no relationship between the Intention-Behavior gap measure and grade point average (GPA).



Scatterplot of all subjects, where I-B gap values are averages of t1 and t2 values when data is available. GPA is calculated out of 100.

However, this was not the case for all measures, as illustrated in Figure @ref(fig:grades-correlation-measures).



Correlations between GPA and various measures. Note: \* p<0.05; \*\* p<0.01; \*\*\* p<0.001

## Out of Sample Prediction

If there exists is a domain-general Intention-Behavior gap, as we propose, we would expect a random sample of the Intention-Behavior gap magnitudes a person had over a number of goal domains to be predictive of an out of sample goal domain gap for that same person. In this case “predictive” means that we believe the expected value of their gap in a held out domain, as calculated based on the average of their gaps in the sampled domains, to contain information about their true gap value in that held out domain. Mathematically this could be expressed as:

Where: - represents the expected value of held out domain . - is the value of the domain. - represents the number of domains in the sampled subset .

We should be clear that while this equation represents the logic of our theory, we do not, in practice, expect to find a precise equivalence, due to individual variance across domains. However, even with this variance we do expect a significantly smaller prediction error if and the sampled s are from the same subject, rather than from a different subject, or taken from the sample population mean. We take three approaches to validating that this is indeed the case.

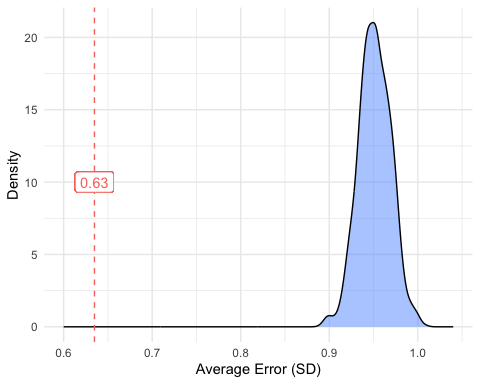
In our first approach, we use a permutation test, where, for a given subject (e.g. ) and specific domain (e.g. ), instead of trying to predict the specific domain gap from the mean of that same subject’s other domain gaps (s), we use a randomly chosen subject’s (e.g. ) average domain gap (excluding the selected domain). We use this methodology to create a set of “predicted” gaps for all of the domains of . We create estimates for all other subjects using this same methodology. We then repeated this process 1000 times and each time calculate the error in each estimate (for a given subject in a given domain ()) as follows:

is the absolute difference between the expected value of (represented by ) and the actual value of . This is then averaged using the following formula:

Where: - represents the total number of subjects. - represents the total number of domains for the subject. - represents the error for the domain of the subject.

The denominator in the fraction to average all of the error scores is instead of simply since each subject can have a different number of goal domains. We then compare the of a matched subject with the distribution of s we get from our permuted subjects to test if a subject’s own goal domain gaps do a better job of predicting a held out domain than a random other subject’s domain gaps. Figure @ref(fig:out-of-sample-gap-prediction) below suggests that this is indeed the case.

To note as well is the fact that we normalized all gap scores before conducting this calculation, given that there are systematic differences in goal domain gap magnitudes (e.g. Volunteering average gap is close to 60% while Work average gap is just below 35%). This means we are trying to predict how many standard deviations above or below the average a subject is in each domain, rather than their actual score.



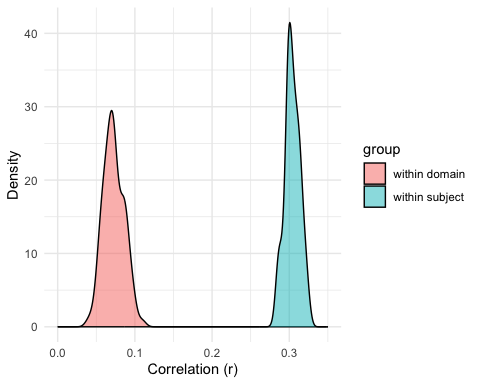
Plot of distribution of average errors. The red dotted line represents the average error when calculated with the actual (non-permuted) subjects data.

A one-sample t-test was computed to determine whether the permuted subject error was different to the within-subject error (0.63). The permuted average prediction error across 1000 repetitions was 0.95 which was significantly higher than the non-permuted within-subject error (95% CI[0.95, 0.952, t(999) = 551.7, p < .001).

The second, alternative approach we take to testing whether there is meaningful signal in our measure used a different prediction approach. In this case we looked to predict a subject’s unmeasured domain gap () by a random sample of another Intention-Behavior gap for that same subject, but in a different domain (e.g. ). This is essentially asserting that for an unknown Intention-Behavior gap () we expect to find a higher correlation between *any* other Intention-Behavior gap domain of that *same subject* and the unknown domain (), than between the unknown domain () and a *different subject’s* gap for that *same* domain (). To represent this mathematically, for a given subject and a randomly selected subject :

Where: - represents a set of domain () values for subject . - is a randomly chosen domain value from . - is a second randomly chosen domain value from (where ). - subject is randomly chosen where ).

To test this we replaced all values for all subjects in all domains (e.g. ) with a randomly chosen other domain from the same subject (e.g. , where ) as well as with a random other subject’s value from the same domain (e.g. , where ). We then tested the average correlation between these predicted values and the actual values. We repeated this process 100 times to come up with a distribution of correlations for both methods, as shown in Figure @ref(fig:out-of-sample-predict-Cendri) below.



Density plots of the correlation of a given domain with either a different domain from the same subject (within subject), or a different randomly selected subject but using the same domain (within domain). The two plots have zero overlap.

The mean correlation in the within-subject calculation was 0.3 (SD = 0.01), whereas the mean for the within-domain calculation was 0.07 (SD = 0.01). A Welch two-samples t-test showed that the correlation for predictions made using a subjects own domains was significantly greater than the correlation between a subject’s gap in a given domain and a randomly chosen other subject’s gap in that same domain (t(180.6) = 141.0416784, p < .001).

Finally, the third approach was to test whether the within-subject method used in the first test of estimating the held out domain () performs better than just taking the domain average for all domains (). We conducted this test with the normalized data for simplicity since in this case all predictions would simply be . We then calculated the average error for both methods. The average error of the within subject method was 0.63 (SD = 0.48) while the average error of the method that just used domain mean values was 0.63 (SD = 0.55). A Welch two-samples t-test showed that the within subject prediction significantly outperformed a sample mean by domain strategy (t(1.0362^{4}) = -19.65, p < .001). Combined, these tests give us confidence that our measure is picking up meaningful signal, and that Intention-Behavior gaps *do* generalize within subject across disparate domains.

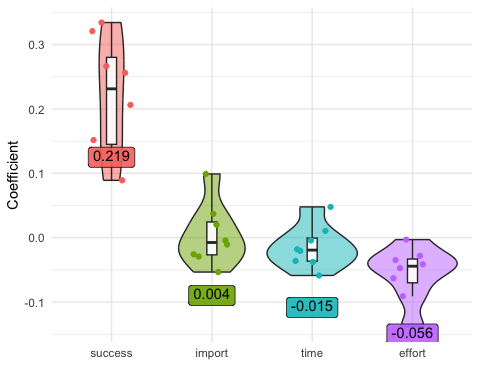
## Informing the Intention-Behavior Gap

As an additional validation we wanted to check whether our Intention-Behavior gap measure had a significant association with our outcome of interest, well-being, above and beyond the hypothesized moderators of the translation of intentions into behavior which we previously found to be correlated with the gap measure itself (e.g. self-control, conscientiousness, etc.). However, before trying to determine models of best fit, we wanted to create a weighted version of the Intention-Behavior gap. This weighed measure would use the additional information we collected about each goal from our subjects.

So far our Intention-Behavior gap measure has ignored the fact that we know more than simply the Intention-Behavior gap in each selected goal domain. Specifically, we collected information about the importance, effort, and time requirements for the goals that our subjects had in each goal domain. Our reasoning is that it seems likely that the degree to which someone’s success in a given goal domain affected their well-being would be related to how important that domain was to them, as well as how much time and effort were required for that goal domain. For example, you might expect those domains that are more important to contribute more to their well-being. To incorporate these additional dimensions to predict our outcome variable we employed a 10-fold cross-validation procedure (to avoid overfitting) using the following equation:

In this equation (Equation @ref(eq:weightedGap-v1)) is the Intention-Behavior gap, is the domain importance, is the effort to accomplish one’s goals in that domain and is the amount of time required to complete one’s goals in that domain. Fitting this model we find that we increase our correlation with our eight well-being outcome variables on average by 6.7%, going from an average correlation of .359 to .383.

The model did not seem to be overfitting, as training and testing error were very similar for the eight outcome variables, with a mean absolute difference between the errors of 0.03 (SD = 0.04), keeping in mind that we are working with normalized data so this difference is in units of standard deviation. It is also interesting to look at the coefficient values for the model. Domain success is clearly the main predictor of our well-being measures as you can observe in @ref(fig:weighted-measure1) below.

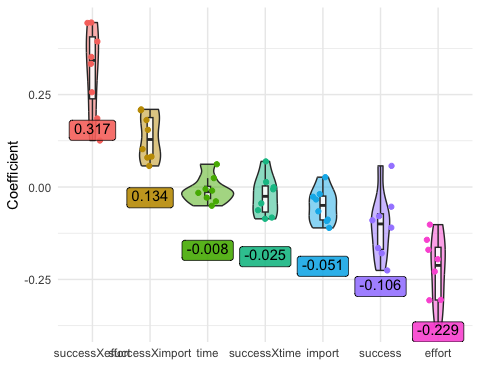


Coefficient violin plots are labeled with mean values across all goal domain dimensions.

However, we could also imagine that the effect of success on well-being isn’t simply a linear, but interacts with our other goal dimensions. For example, you might imagine that the a given level of success on a very important goal would have a greater impact on well-being than the same level of success on a goal of only modest importance. We have therefore constructed a second weighted-gap model that includes interactions between the Intention-Behavior gap and importance, effort, and time, as shown by the three terms added to the previous equation. We again used 10-fold cross-validation to fit the model.

Here (Equation @ref(eq:weightedGap-v2)), for example, is the interaction, for a given domain, between a subject’s gap and the importance of that gap ( represents effort and , time). This model correlated even more strongly with our well-being outcome variables, with a 19.8% improvement in average correlation values (from to ) over the “gap only” model, and a 12.3% increase over the weighted model without interaction terms (from to ).

Again, the model did not seem to be overfitting as training and testing error were very similar for the eight outcome variables, with a mean absolute difference of 0.03 (SD = 0.04, units are standard deviation).



Coefficient violin plots are labeled with mean values across all goal domain dimensions, and interactions.

Looking at the coefficient values for the model a few interesting observations can be immediately made. First, domain success is no longer the main predictor of our well-being measures - and not only that, but the sign of the success coefficient has flipped, from positive (, p<??) to negative (). This raises the interesting idea that greater than average success in accomplishing a goal that has only average importance, takes an average amount of time, and requires an average effort, actually has a negative impact on well-being (recall that this data is normalized and so the sample average maps to zero). Less surprisingly the model posits that if you achieve only average success on a goal that expending above average effort in achieving that goal lowers well-being (). The other most predictive variables in the model were the interactions between the gap and effort (), as well as the gap and importance (). The fact that the interaction between success and effort was the single best predictor surprised us as we had expected the interaction between success and importance to be more important.

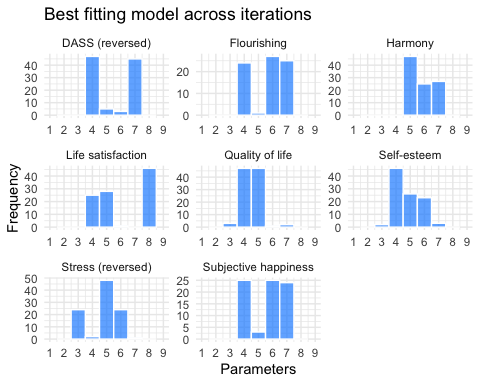
## Model Selection

We wanted to test whether our Intention-Behavior gap measure explained unique variance in our well-being outcome measures when other predictors were included in the same model.

### Individual Outcome Variable Prediction

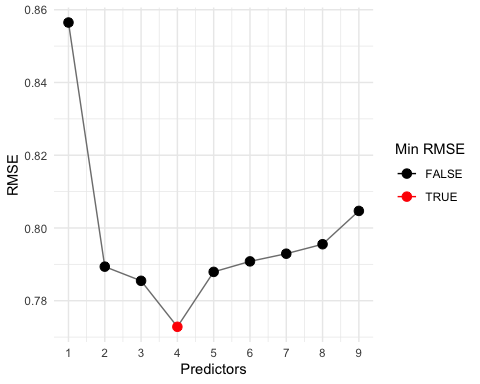
Specifically, we added all moderator variables and the weighted Intention-Behavior gap then iterated through all possible models using the leaps package (v3.1; Lumley, 2020). This selects the best fitting model for a given dependent variable (in this case our outcome measures) for each possible number of predictor variables. In our case given that we had nine predictors (our eight moderator variables plus our weighted Intention-Behavior gap measure), we ended up with a series of models, starting with one predictor and going all the way up to nine predictor variables.

We then evaluated each of these models individually using nested (100 repetitions) five-fold cross-validation, to avoid overfitting, with the caret package (version 6.0.94; Kuhn, Max, 2008). Each repetition of the cross-validation process selected a best fitting model for each outcome measure, based on minimizing RMSE on the held out measure across folds. Across iterations of this process the model with the lowest RMSE varied, usually fluctuating between two or three models that best predicted a given outcome (Figure @ref(fig:measure-fit-comparisons)). The fact that the best-fitting model varied in repeated iterations of the cross-fold validation indicates that there were multiple models that had very similar fits, and the random splitting of the data led to one being selected over another in a given iteration. We selected the modal model - that is to say the model that was most often selected as best fitting over the 100 iterations - as the final best fitting model.



Bars show frequency that a model with a given number of parameters was selected in the cross-validation process. In total 100 iterations of the cross-fold validation process were run.

Figure @ref(fig:model-selection-self-esteem) shows what this cross-fold validation process looked like for the Rosenberg Self-Esteem Scale (one of our eight well-being outcome measures) for a single iteration. Here the dots represent the average of the RMSE across the five folds, and we can observe that the lowest RMSE was acheived in the model with four independent variables. Looking back at Figure @ref(fig:measure-fit-comparisons) we can see that this was the modal model, although there were a number of iterations for which different models achieved best fit (e.g. five and six parameter models).



Average RMSE (using 5-fold cross validation) for models with various numbers of independent variables predicting the self-esteem outcome.

In Table @ref(tab:model-selection-table) we can see which variables are included in this 4-predictor model for self-esteem. Similarly, for all other outcome measures we present the model of best fit along with the standardized beta coefficient, averaged across all models, and the 95% confidence interval of the beta coefficient. A grey background indicates that the variable is not included in the selected model of best fit.

Predictor column cells show beta coefficients with 95% confidence intervals. The coefficients are the mean of the coefficient value across all iterations of the nested cross-validation process. Coloring corresponds to the magnitude of the beta coefficient. The cells with a ’\*’ (colored grey) were not included in the final best fitting model.

Outcome

RMSE

DF

R\_2

Intention-behavior gap

Self-control

Conscientiousness

Social desirability

Future-perspective

Work ethic

Sensation-seeking

Grit

Ambition

DASS (reversed)

0.88

236

0.23

-0.24 [-0.24, -0.24]

0.33 [0.32, 0.34]

-0.15 [-0.15, -0.15]

0.07 [0.05, 0.08]\*

-0.02 [-0.02, -0.01]\*

0.01 [0.01, 0.02]\*

-0.14 [-0.15, -0.14]

0.00 [NaN, NaN]\*

0.00 [NaN, NaN]\*

Flourishing

0.77

234

0.43

-0.31 [-0.31, -0.30]

0.15 [0.14, 0.16]

0.01 [0.01, 0.01]\*

0.06 [0.05, 0.07]

0.27 [0.27, 0.27]

0.14 [0.14, 0.15]

0.00 [0.00, 0.00]\*

0.08 [0.07, 0.09]

0.00 [0.00, 0.00]\*

Harmony

0.82

235

0.34

-0.33 [-0.34, -0.33]

0.28 [0.27, 0.28]

-0.24 [-0.24, -0.24]

0.12 [0.12, 0.12]

0.06 [0.05, 0.07]\*

0.17 [0.16, 0.17]

-0.02 [-0.02, -0.01]\*

0.00 [-0.00, 0.00]\*

-0.00 [-0.00, 0.00]\*

Life satisfaction

0.83

232

0.35

-0.28 [-0.28, -0.28]

0.36 [0.36, 0.37]

-0.21 [-0.21, -0.21]

0.03 [0.03, 0.04]

0.17 [0.17, 0.17]

0.07 [0.06, 0.08]

0.00 [NaN, NaN]\*

0.03 [0.02, 0.03]

-0.03 [-0.03, -0.02]

Quality of life

0.88

236

0.23

-0.34 [-0.34, -0.34]

0.00 [-0.00, 0.00]\*

0.00 [NaN, NaN]\*

0.14 [0.14, 0.14]

0.00 [NaN, NaN]\*

0.14 [0.13, 0.14]

0.04 [0.04, 0.05]\*

0.05 [0.04, 0.06]\*

0.05 [0.04, 0.06]

Self-esteem

0.77

236

0.40

-0.34 [-0.34, -0.34]

0.34 [0.34, 0.35]

-0.05 [-0.06, -0.04]\*

0.10 [0.10, 0.11]

0.12 [0.12, 0.13]

0.02 [0.01, 0.03]\*

0.00 [-0.00, 0.00]\*

0.00 [NaN, NaN]\*

0.00 [NaN, NaN]\*

Stress (reversed)

0.82

235

0.35

-0.29 [-0.29, -0.29]

0.42 [0.42, 0.43]

-0.20 [-0.20, -0.20]

0.08 [0.07, 0.09]

-0.02 [-0.03, -0.01]\*

-0.00 [-0.00, 0.00]\*

-0.08 [-0.09, -0.07]

0.00 [-0.00, 0.00]\*

-0.00 [-0.00, 0.00]\*

Subjective happiness

0.83

236

0.32

-0.35 [-0.35, -0.35]

0.25 [0.25, 0.25]

-0.28 [-0.29, -0.28]

0.16 [0.16, 0.16]

0.05 [0.05, 0.06]\*

0.08 [0.07, 0.09]\*

0.01 [0.00, 0.01]\*

0.01 [0.01, 0.01]\*

-0.02 [-0.03, -0.02]\*

MEANS

0.83

235

0.33

-0.31 [-0.31, -0.31]

0.27 [0.26, 0.28]

-0.14 [-0.15, -0.13]

0.10 [0.09, 0.10]

0.08 [0.07, 0.09]

0.08 [0.07, 0.08]

-0.02 [-0.03, -0.02]

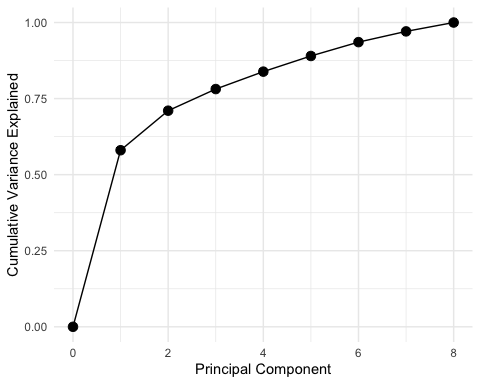
0.02 [0.02, 0.02]

0.00 [-0.00, 0.00]

Table @ref(tab:model-selection-table) also shows that the weighted Intention-Behavior gap measure was the only variable that was included in every model for all of our eight well-being outcome measures. The bottom row of the table shows the average absolute magnitude of each feature’s standardized beta weights in the best-fitting models. This average was calculated using for a feature’s beta weight if it was not included in the best fitting model. Again, the Intention-Behavior gap appears to have the highest average standardized regression coefficient () of .31, although the difference between this value and that of Self-control () is not significant () according to a Welch two sample t-test. The difference between the Intention-Behavior gap measure and average magnitude of the Conscientiousness coefficient (), however, is highly significant (), as are the differences with the rest of the predictor variables.

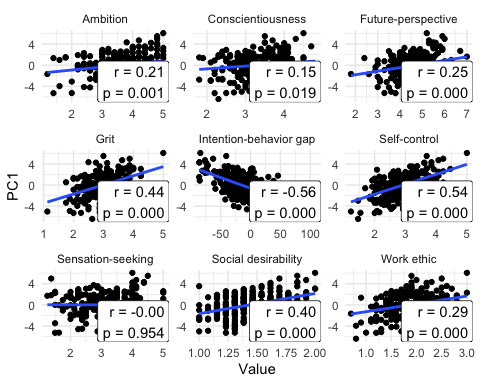
### Combined Outcome Variables Prediction

An alternative approach is to combine the eight outcome variables into a single value, and then to test how predictive each of our designated moderators and the Intention-Behavior gap measure are of this combined outcome value. We perform this amalgamation of the eight well-being outcome variables using principal components analysis. As shown in the scree plot (Figure @ref(fig:pca-outcomes-moderators)) there is a pronounced “elbow” at the first component, indicating that the bulk of the variance is explained by this single component on its own.



Scree plot of cumulative variance explained with each additional principal component.

The first component successfully accounts for over half (58%) of the total variance of the outcome variables. Given that this single components seemed like a meaningful combination of our eight outcome variables we looked at the correlations between this first component and our moderator variables and the weighted Intention-Behavior gap measure (Figure @ref(fig:pca-outcomes-moderators-plot)).



Correlation plots of the Intention-Behavior gap measure and moderator variables with the first principle compoment of all outcome measures.

The strongest correlation was between the well-being outcomes’ first principle component and the weighted Intention-Behavior gap measure (r(246) = -0.56, 95% CI [-0.64, -0.47]). However, it was similar in magnitude and had an overlapping 95% confidence interval, in absolute terms, with self-control (95% CI [0.44, 0.62]), grit (95% CI [0.33, 0.53]) and social desirability (95% CI [0.29, 0.5], all s ).

We calculated the best fitting model (using 10-fold cross validation) of the well-being outcomes’ first principle component (Table @ref(tab:pca-outcomes-moderators-model)). We again found the Intention-Behavior gap measure and self-control to have strong effects () on well-being with moderate effects () of conscientiousness, social desirability, work ethic and future time perspective measures. To assess the potential for multicollinearity among the predictor variables, variance inflation factors (VIFs) were computed. All VIF values were below the commonly used threshold of 5 (all ), suggesting that multicollinearity is not a concern in the current model. Interestingly, grit, which correlated strongly on its own with the outcomes’ first principle component did not have a significant unique contribution to the final model.

Best model fit to outcome measures first principal component (standardized beta weights)

RMSE

DF

R\_2

Intention-behavior gap

Self-control

Conscientiousness

Social desirability

Work ethic

Future-perspective

Ambition

Sensation-seeking

Grit

0.72

234

0.52

-0.4

0.37

-0.19

0.15

0.12

0.11

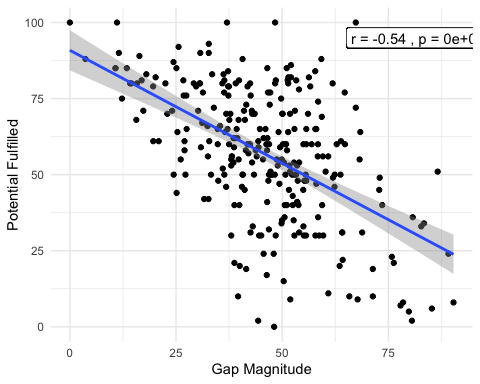
NA

NA

NA

## The Gap and Individual Potential

Given that it is possible to have an Intention-Behavior gap even while working as hard and efficiently as one possibly can (e.g. just set impossibly ambitious goals), we wanted to confirm our intuition that this theoretically possible situation is an edge case rather than a general explanation of the Intention-Behavior gap. To test this we added a question at the end of the measure asking subjects to indicate to what degree they felt like they were currently fulfilling their potential abilities (0-100%). The idea being that if you are currently fulfilling 100% of your potential then any gap you may have between intentions and behavior would best be solved by just setting more realistic goals, since you are already at ceiling in terms of goal accomplishment.



As expected (Figure @ref(fig:gap-versus-potential)), there was a strong negative association between gap magnitude and self-reported fulfillment of one’s personal potential (r(280) = -0.54, p < .001). We interpret this as strong support for the idea that the Intention-Behavior gap is not solely driven by ambitious goals, but also, at least in part, by behavior that is un-aligned with accomplishing one’s goals.

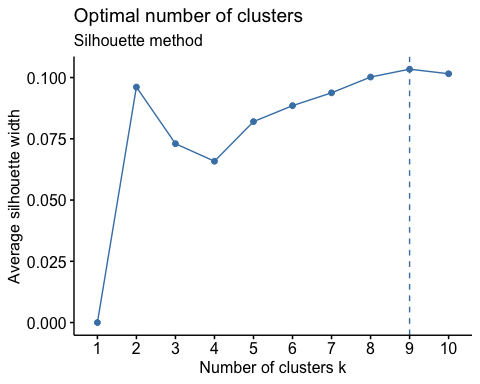
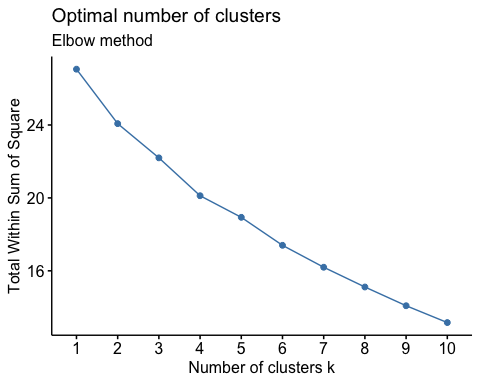
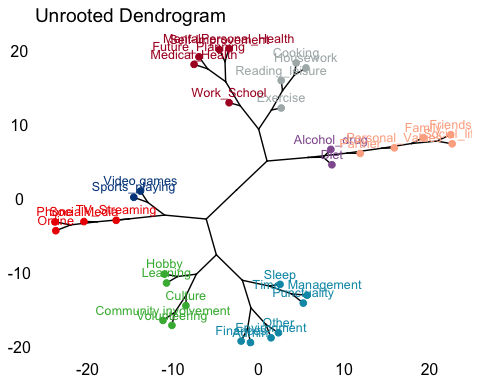
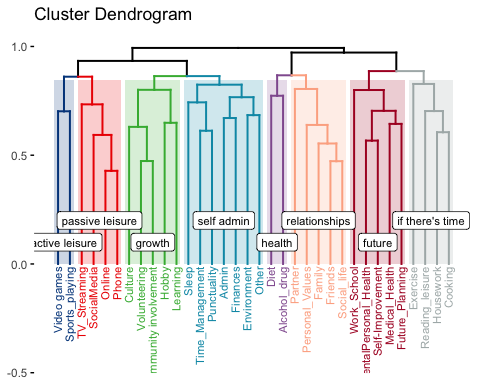
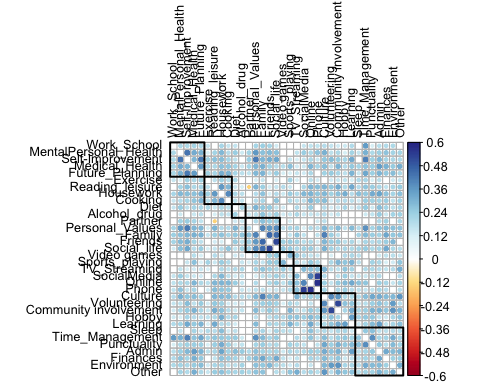
# Exploratory Analyses

We had many additional questions that we wanted to shed light on with the data collected for the Intention-Behavior gap measure. To start we look at whether there is interpretable structure contained within the reported goals. Do goal domains group together in ways that we would expect? If not what might that suggest? We also collected general information from subjects about their general tendencies when it came to setting and persuing their goals. We were interested to see in what ways these characteristics might be predictive of the Intention-Behavior gap. For example, while we might expect having a strong ability to prioritize goals might reduce your Intention-Behavior gap, is this intuition actually supported by the data? This also connects to another question we had regarding goal conflict. It is clear that goal pursuit is often a zero-sum game. If you are engaged in your goal of playing sports you cannot simultaneously be attending to your goal of cleaning the apartment. Given that time is a finite resource these types of tensions can be expected to often arise between different goals. However, there are also cases in which you could imagine goals actually symbiotically reinforcing each other. For example your goal to cook your own meals might actually help with your goal to follow a particular diet you have set for yourself. We look into a number of these questions in the [Goal Conflict](#goal-conflict) section. Given our fairly modest test-retest values for our measure (see [Test-Retest Reliability](#test-retest-reliability)) we were curious whether some peolple might simple be more variable than others when it came to their Intention-Behavior gap. We tested to see if variability in daily reports of subjects’ Intention-Behavior gap was associated with higher variability in their t1 and t2 gap measures. Finally, we looked briefly at a couple of different ways in which gender differences in goal domains manifested.

## Exploratory Factor Analysis

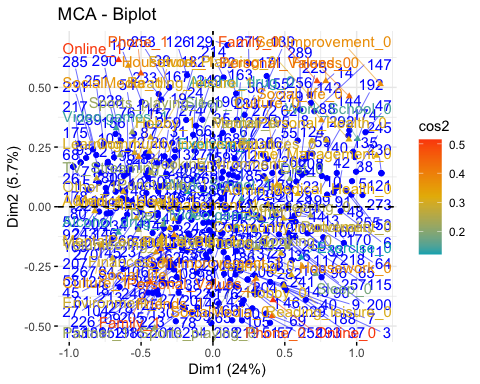
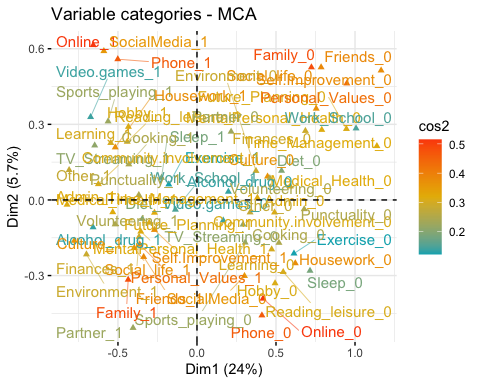
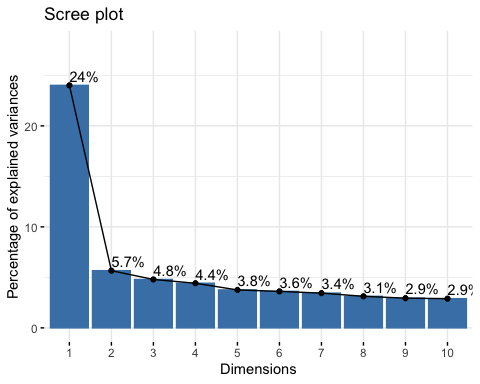
### Goal Domain Structure

Given that we have people reporting the specific life domains in which they have goals we can look to see whether there might be interesting structure in the data where certain goal domains might group together. The groupings could suggest that a set of domains have common characteristics. For example, if a subject has a goal in domain ‘A’, will they be likely to also have a goal in domain ‘B’? To do this we create a correlation table of bi-serial correlations between each pair of goal domains. We can then use hierarchical agglomerative clustering to create groupings.



Using this methodology we decided on eight groups that seemed relatively interpretable: active and passive leisure, personal growth, self-administration, health, relationships, future, and “if there’s time”.

We can also use the presence or absence of goals in each domain as a means of grouping subjects. We use multiple correspondence analysis (MCA, close relation of PCA) since we are dealing with categorical data (had a goal in a domain or did not).

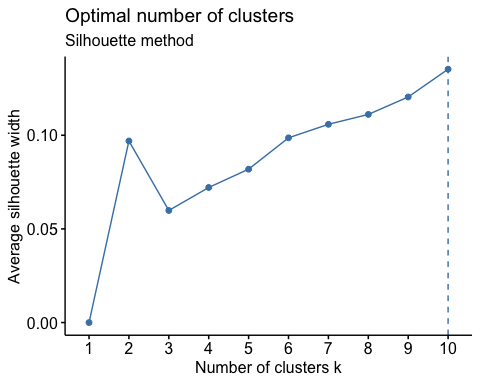
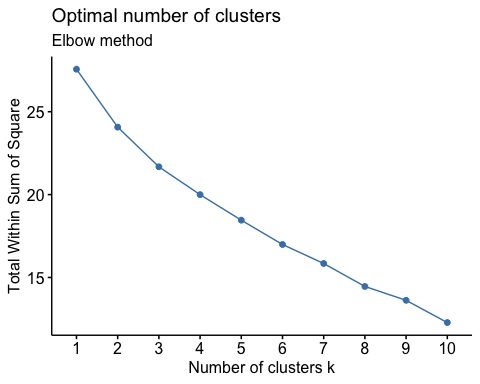
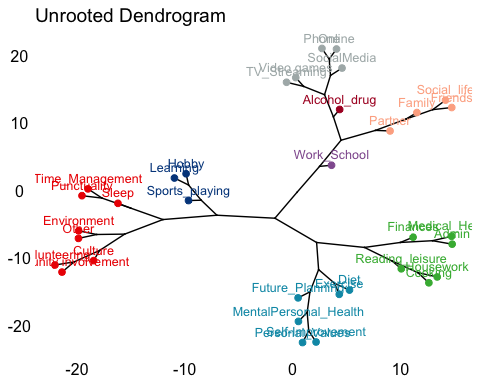
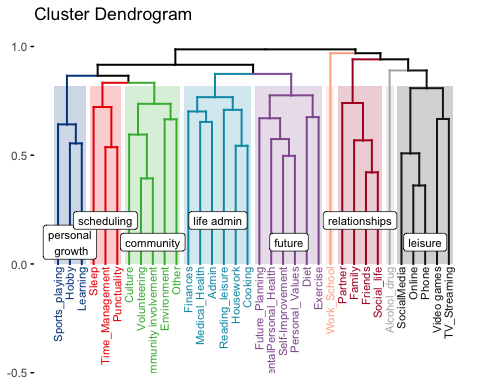
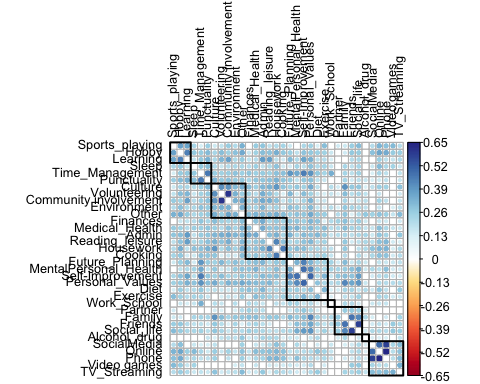


## # A tibble: 4 × 5  
## quadrant mean\_gap mean\_grades mean\_happy n  
## <dbl> <dbl> <dbl> <dbl> <int>  
## 1 1 48.3 74.0 4.14 76  
## 2 2 46.0 76.3 4.27 73  
## 3 3 44.8 75.2 4.1 66  
## 4 4 44.4 73.7 4.21 78

We did not find any significant differences between subjects in the four different quadrants, so this may not be a particularly meaningful grouping.

### Goal Domain Importance

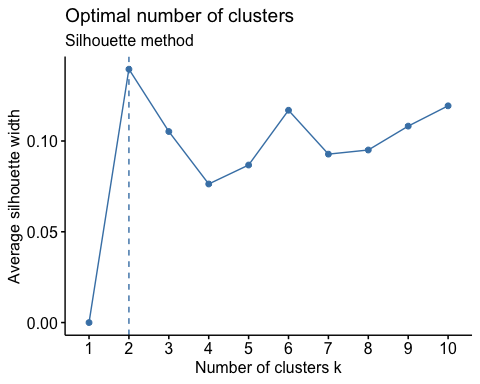
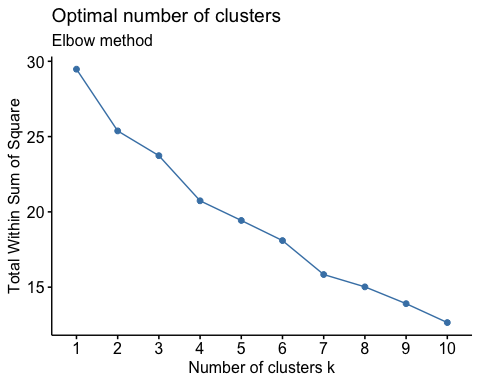
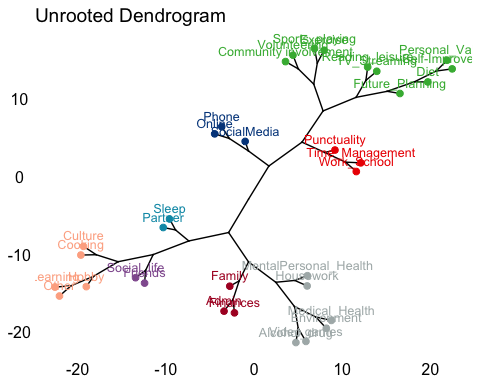
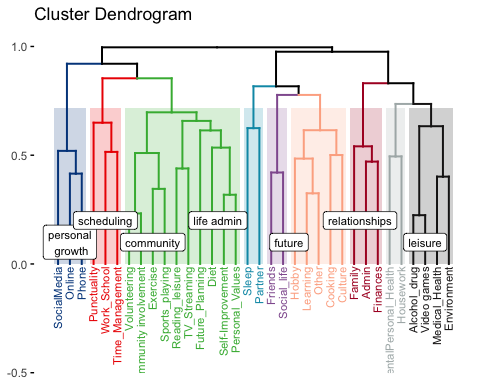
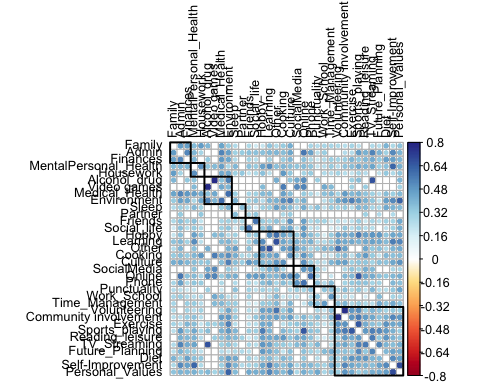
We can do a similar analysis of goal domains based on importance, rather than presence/absence of a goal. For example, if domain A is of high importance does this correlate with domain B also being highly important?



In this case we had a different grouping than for goal presence/absence. For example, instead of playing sports being grouped with video games, it is now clustered with “hobby” and “learning”, a grouping we have labeled “personal growth”. Work/School and Alcohol/Drug also form their own solo clusters in this grouping.

### Goal Domain Success

Finally, we conducted a similar analysis this time looking at domain success (the flip side of the gap). If you achieve success in domain A are you likely to also achieve success in domain B?

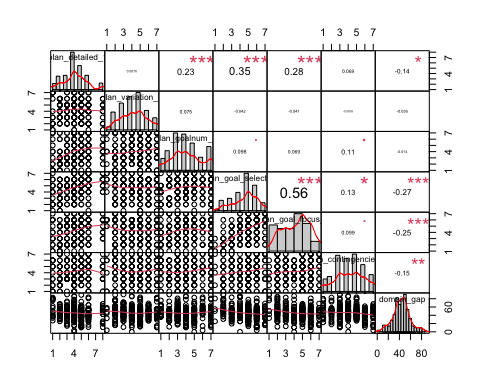


## Goal Characteristics and the Intention-Behavior Gap

### Goal Pursuit Characteristics

We asked people a number of questions about their general goal setting style to try and gain some insight into which strategies or approaches might be more associated with the Intention-Behavior gap. These included: - How detailed are your goals? - How variable are your goals on a day to day basis? - How many goals do you tend to have? - How good are you at selecting a specific goal to pursue at any given time? - How easy is it for you to focus on one specific goal? - Do you tend to plan for contingencies with your goals?

See the plot below to observe how these correlate with the (unweighted) Intention-Behavior gap.



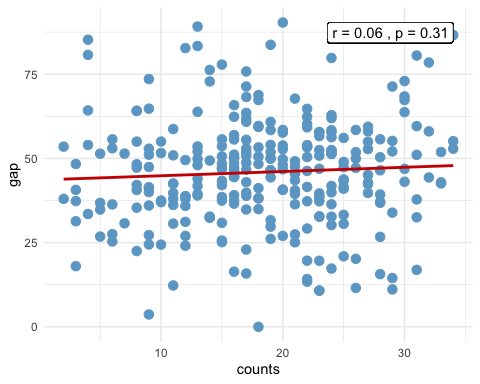
##   
## Call:  
## lm(formula = domain\_gap ~ plan\_goal\_select\_1 + plan\_goal\_focus\_1 +   
## plan\_goalnum\_1 + plan\_contingencies\_1 + plan\_variation\_1 +   
## plan\_detailed\_1, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -48.107 -8.408 0.200 9.322 39.982   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 67.2837 4.8573 13.852 <2e-16 \*\*\*  
## plan\_goal\_select\_1 -1.7964 0.7812 -2.299 0.0222 \*   
## plan\_goal\_focus\_1 -1.3579 0.6772 -2.005 0.0459 \*   
## plan\_goalnum\_1 0.3619 0.5300 0.683 0.4953   
## plan\_contingencies\_1 -1.1248 0.5416 -2.077 0.0388 \*   
## plan\_variation\_1 -0.5037 0.6021 -0.837 0.4036   
## plan\_detailed\_1 -0.4638 0.6154 -0.754 0.4518   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 15.05 on 275 degrees of freedom  
## (45 observations deleted due to missingness)  
## Multiple R-squared: 0.1041, Adjusted R-squared: 0.08459   
## F-statistic: 5.328 on 6 and 275 DF, p-value: 3.215e-05

##   
## Call:  
## lm(formula = domain\_gap ~ plan\_goal\_select\_1 + plan\_goal\_focus\_1 +   
## plan\_contingencies\_1, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -47.625 -8.782 0.286 9.828 38.993   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 65.1595 3.6838 17.688 <2e-16 \*\*\*  
## plan\_goal\_select\_1 -1.8920 0.7567 -2.500 0.0130 \*   
## plan\_goal\_focus\_1 -1.3961 0.6717 -2.079 0.0386 \*   
## plan\_contingencies\_1 -1.0956 0.5374 -2.039 0.0424 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 15.01 on 278 degrees of freedom  
## (45 observations deleted due to missingness)  
## Multiple R-squared: 0.09917, Adjusted R-squared: 0.08944   
## F-statistic: 10.2 on 3 and 278 DF, p-value: 2.15e-06

If we include all of these characteristics in a model with the Intention-Behavior gap as the dependent variable, we observe that only the ability to prioritize goals for pursuit, the ability to focus on a single goal at a time, and the tendency to plan for contingencies remain significant predictors of the gap (ps < .05).

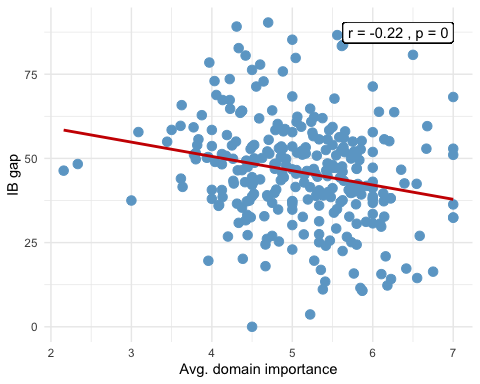
### Number of goals

While we were not able to check on each of these goal setting style attributes in our measure data, we were able to at least test whether the number of goals was, in fact, not significantly related to the gap given that we had reports of both quantities for our participants. This was indeed the case with a non-significant relationship (r(286) = .06, p = .31).



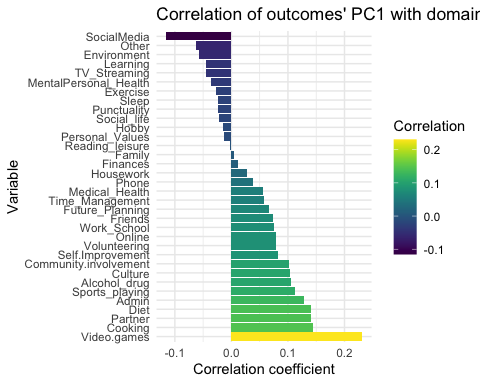
### Goal importance

Does having on average more important goals mean you are more likely to accomplish them?



## Domain Importance for Outcomes

Using the first principle component of our eight well-being outcome measures, do certain goal domains have a higher association with well-being than others?



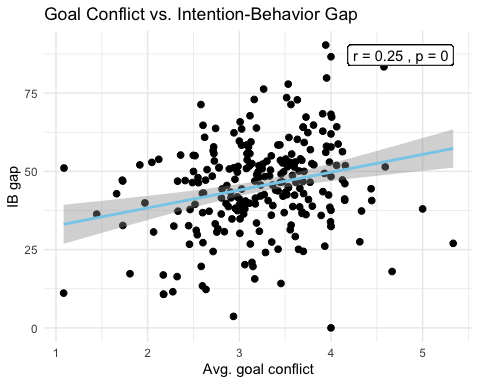
## Goal Conflict

Given that it is often the case that goal pursuit is a zero-sum scenario in which time spent pursuing goal A is taking away from time available to pursue goal B, it is clear that conflict between and among goals can arise.

We attempted to quantify this conflict for our subjects by taking their top seven goal domains, and then for each of these domains specifying whether their goal in the selected domain promoted or conflicted with their goals in each of the other six domains (or was independent).

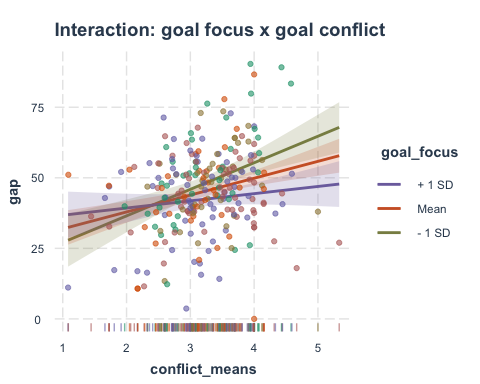
### Goal Conflict and the Intention-Behavior Gap

We first tested our basic intuition that higher average goal-conflict level for a subject would predict a larger Intention-Behavior Gap measure.

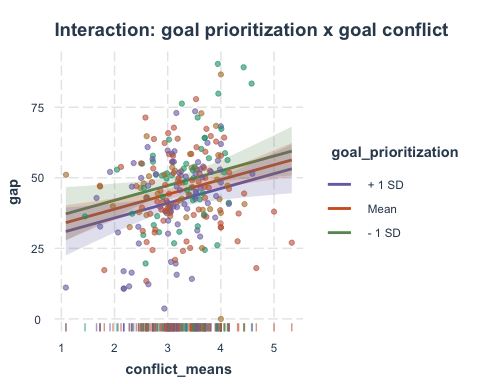


We found this to be the case with a positive correlation of r(246) = .25 (p < .001). Given that we had self report information on how well subjects could prioritize and focus on goals, we were also intersted to test whether these goal strategies might moderate the relationship between goal conflict and the intention behavior gap.

##   
## =========================================================================  
## Dependent variable:   
## -----------------------------------------------  
## gap   
## Main Effects Interaction   
## (1) (2)   
## -------------------------------------------------------------------------  
## Constant 35.323\*\*\* 3.455   
## p = 0.000 p = 0.808   
##   
## conflict\_means 5.544\*\*\* 15.326\*\*\*   
## p = 0.0001 p = 0.0004   
##   
## goal\_focus -1.826\*\* 5.192   
## p = 0.002 p = 0.080   
##   
## conflict\_means:goal\_focus -2.159\*   
## p = 0.017   
##   
## -------------------------------------------------------------------------  
## Observations 248 248   
## R2 0.099 0.120   
## Adjusted R2 0.092 0.110   
## Residual Std. Error 14.336 (df = 245) 14.196 (df = 244)   
## F Statistic 13.505\*\*\* (df = 2; 245) 11.141\*\*\* (df = 3; 244)  
## =========================================================================  
## Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

 This was, in fact, what we found. If we included an interaction between a subject’s ability to focus on a specific goal and their average goal conflict level, then this relationship between the magnitude of conflict and magnitude of gap is reduced. We tested this also for goal prioritization but did not find the same relationship.

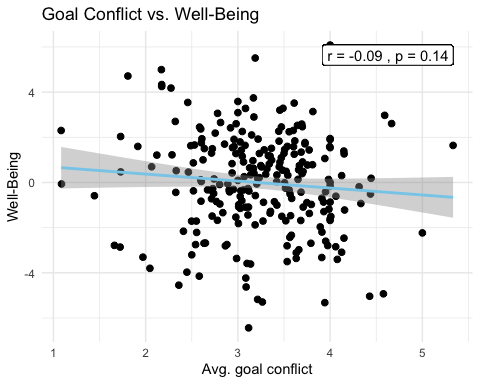
##   
## =================================================================================  
## Dependent variable:   
## ----------------------------------------------  
## gap   
## Main Effects Interaction   
## (1) (2)   
## ---------------------------------------------------------------------------------  
## Constant 39.089\*\*\* 39.050\*   
## p = 0.000 p = 0.028   
##   
## conflict\_means 5.230\*\*\* 5.242   
## p = 0.0003 p = 0.318   
##   
## goal\_prioritization -2.275\*\*\* -2.267   
## p = 0.001 p = 0.521   
##   
## conflict\_means:goal\_prioritization -0.002   
## p = 0.999   
##   
## ---------------------------------------------------------------------------------  
## Observations 248 248   
## R2 0.105 0.105   
## Adjusted R2 0.098 0.094   
## Residual Std. Error 14.291 (df = 245) 14.320 (df = 244)   
## F Statistic 14.377\*\*\* (df = 2; 245) 9.545\*\*\* (df = 3; 244)  
## =================================================================================  
## Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001



In this case it appeared that there was no interaction at all, with goal prioritization and goal conflict having independent effects on the gap.

### Goal Conflict and Well-being

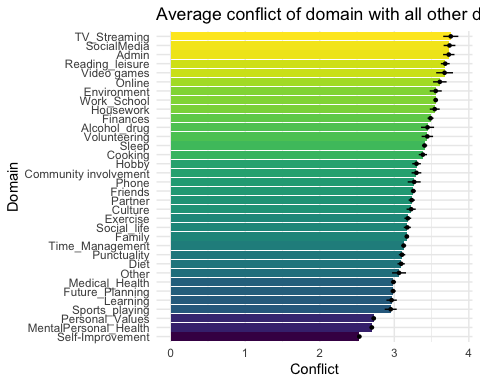
We were also interested in whether there would be a negative relationship between average goal conflict and well-being. Our intuition was that having greater goal conflict in life might lead to a reduction in well-being (as quantified by the first principal component of our eight well-being measures).



While the relationship was negative, it was not significant (r(246) = -0.09, p = 0.14).

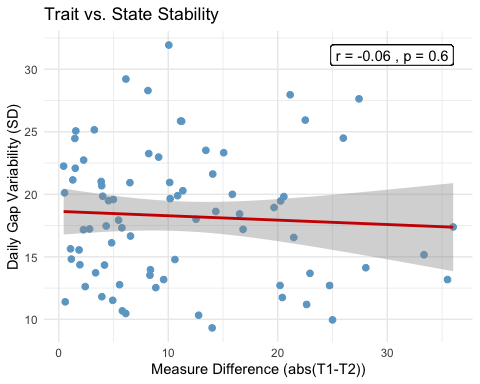
### Average Goal Conflict by Domain

Finally, we looked at the average level of conflict by domain. Perhaps unsurprisingly watching television (streaming) and social media were the two domains that were highest in conflict with all other domains.



## Trait vs State Stability

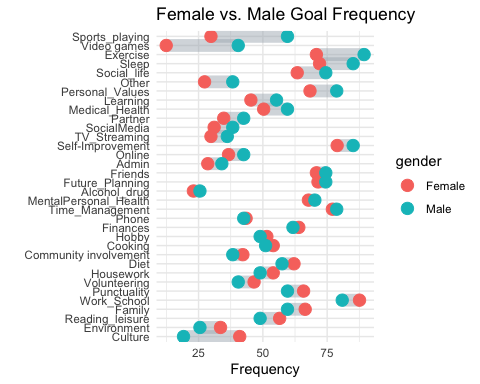
Given that we have modest test-retest reliability it is clear that subjects’ self-reported Intention-Behavior gap often differed between the two timepoints where measurements were collected. We were interested to see if the magnitude of a subject’s change might be associated with their self-reported gap stability in the daily gap state measurements we collected. Our intuition was that someone who had higher day to day variability, as assessed via the standard deviation of their score, might tend to show larger magnitude changes in their trait level measure from collection timepoint one to two.

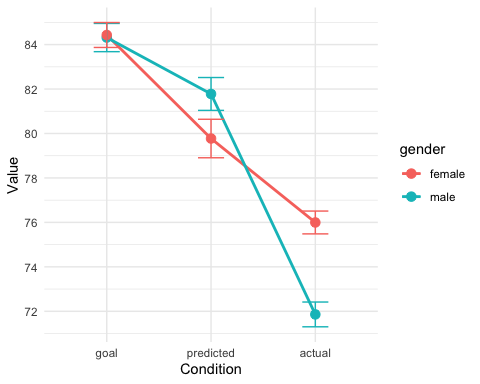


We correlated the magnitude of the Intention-Behavior gap difference between timepoint one and two, and the state-level (daily) Intention-Behavior gap measure variability (SD). We did not find a significant correlation ().

## Gender Differences

We also noticed some interesting and thought-provoking trends when we separated subjects be gender. First, the frequency with which different domain goals were selected varied substantially, though not for all domains. The domains for which males had the largest difference in terms of the frequency with which they selected goals as compared to females included playing sports, video games, exercise and sleep, whereas for females the domains in which they were more frequently reporting goals as compared to males included culture, environment, reading for leisure and family.

 We also found a notable trend in which males tended to overestimate their future academic success much more than females.

 While males and females did not significantly differ in terms of reporting their ‘goal’ grades, males ‘predicted’ grades (what they thought they would actually get) tended to be higher (M = 81.8, SD = 8.65) than females (M = 79.8, SD = 10.14), though not significantly (p = 0.28). However, when it came to actual grades (as collected from the registrar) things flipped, with males grades now significantly lower than females (t(71.2) = -2.6, p = .01).