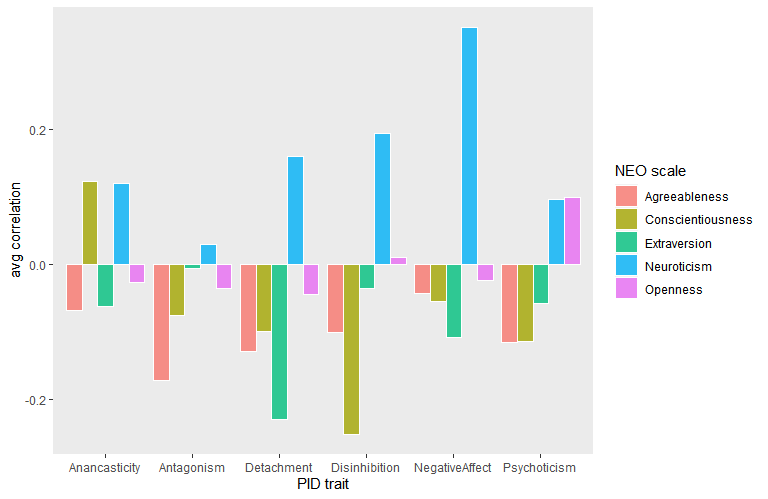
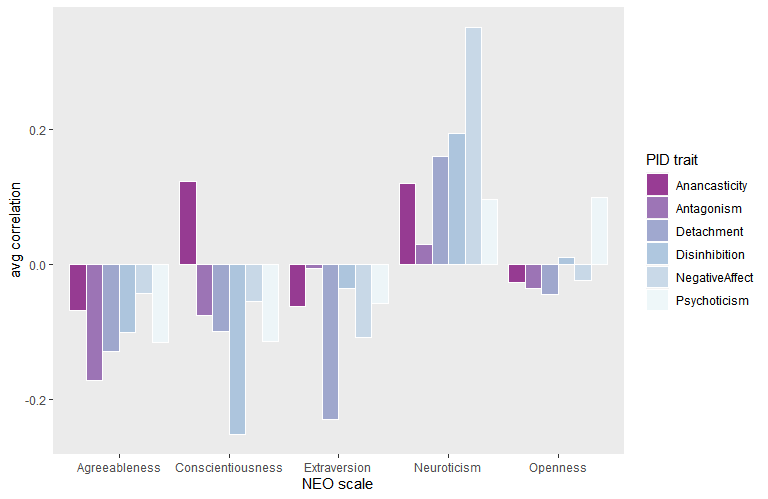
Ontology Models

Roberto Viviani, University of Innsbruck

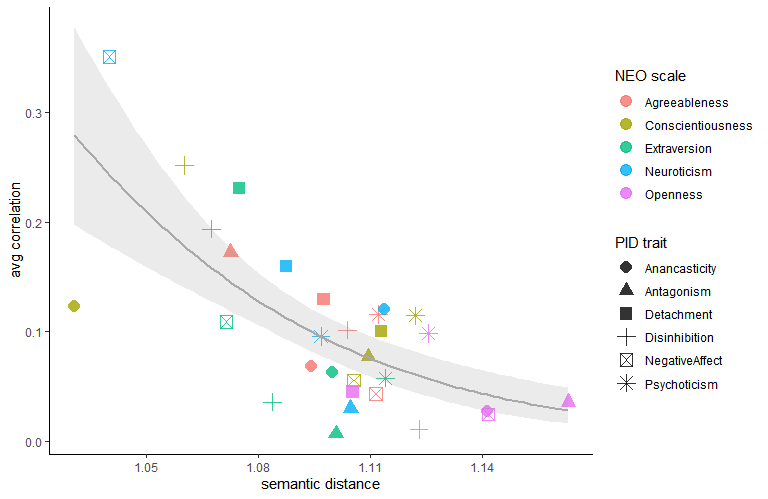
2024-04-18

The question that naturally arises is the extent to which semantic similarity explains the correlation between scales. The average correlation between NEO and PID scales, computed from the correlation of the items, is shown in the following figures.



We see here the pattern of a positive association between neuroticism and PID scores, and a negative association of all other NEO scales, with the exception of openness to experience that had no association, except to psychoticism. Conscientiousness is another exception to this pattern because of the diverging associations with anancasticity (positive) and disinhibition (negative).

To see how much these correlations are explained by semantic relatedness, we plotted their absolute value on semantic relatedness (we used the absolute value to redress the issue of the differing polarity of the NEO subscales).

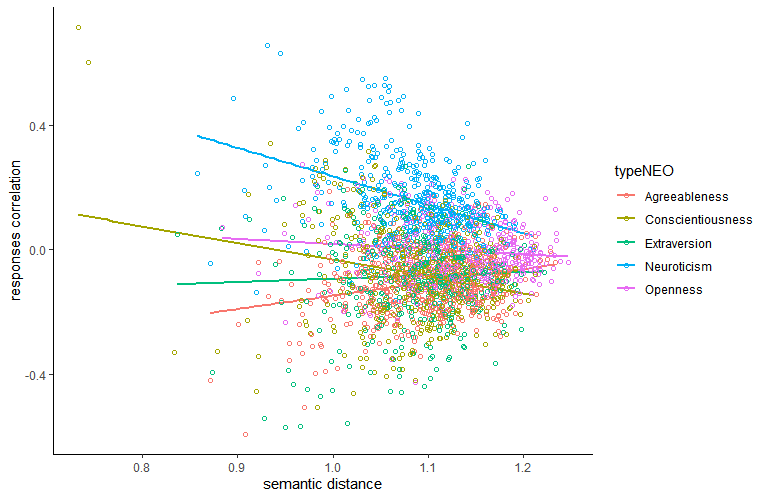


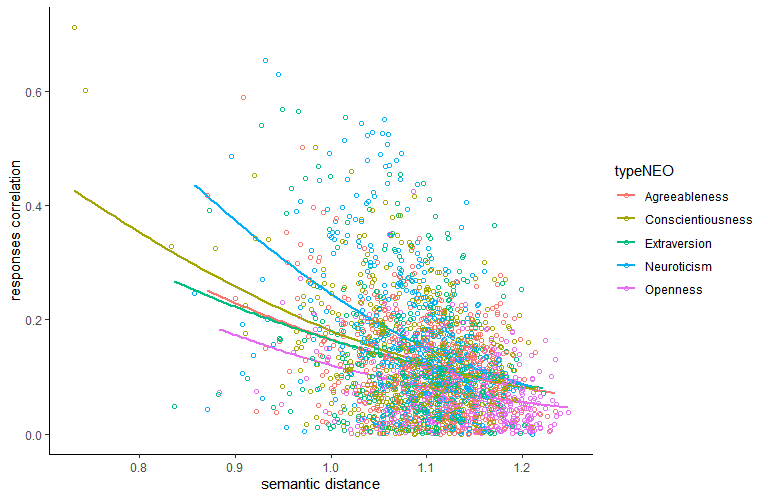
##   
## Call:  
## glm(formula = abs(corr) ~ dist, family = quasibinomial, data = summarize(group\_by(cordata,   
## typeNEO, traitPID), corr = mean(corrs), dist = mean(cosdist)))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.37802 -0.11443 0.00168 0.13497 0.24476   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 19.366 3.928 4.931 3.35e-05 \*\*\*  
## dist -19.708 3.611 -5.458 7.95e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for quasibinomial family taken to be 0.02977943)  
##   
## Null deviance: 1.81380 on 29 degrees of freedom  
## Residual deviance: 0.92642 on 28 degrees of freedom  
## AIC: NA  
##   
## Number of Fisher Scoring iterations: 5

In general, the closer the semantic relatedness of items of the respective subscales and traits, the stronger the correlations in the answers of participants (t = -5.5, p < 0.001). We also see that the strongest correlation, given by neuroticism and negative affect, was also accompanied by a short semantic distance between the items of the respective scales. However, the amount of correlation is somewhat higher than we would expect from a linear prediction from the amount of semantic relatedness. We can see also that conscientiousness departed from this pattern. However, the semantic construct implied by the conscientiousness items appears to be heterogeneous. As described in the literature, openness to experience was not related to PID dimensions. We see that in both its large semantic distance and low correlation with total PID scores.

## Association of individual items in the NEO subscales

In the following plots, we display the correlation of pairs of individual items on the cosine distance of embeddings to visualize the relation between semantic relatedness and correlation at a smaller level of granualarity. We use color to differentiate the PID facets, and plot the relationship between pairs with curves grouped by NEO subscales (which have different directions).





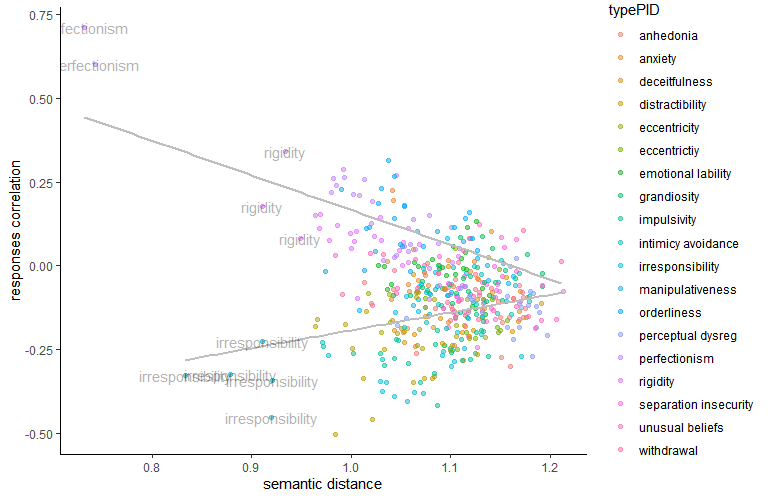
To interpret these plots it is important to remember that each point represents a pair of scale items, one from the NEO inventory and one from the PID scale, from which we take the correlation of responses and the semantic distance. Therefore, as there are 66 NEO items in this database and 36 PID items, there are 66 x 36 = 2376 points in total.

The first plot demonstrates the diverging relationship between semantic distance and response correlations. At low semantic distance, correlations diverged widely due in part to the fact that some have opposite directions. At high semantic distances, in contrast, we see that the correlations narrowed around the zero point.

The second plot considers the absolute values of the correlations, to eliminate the obfuscating role of scale items with similar meaning but opposite polarity. This plot shows the extent to which low semantic distance was associated with varying degrees of association in the participant responses. From the fitted responses (obtained with a logistic binomial regression), however, it is clear that on average correlation increased with low semantic distance. Furthermore, we see that neuroticism displayed the strongest association with semantic similarity, and openness to experience the lowest.

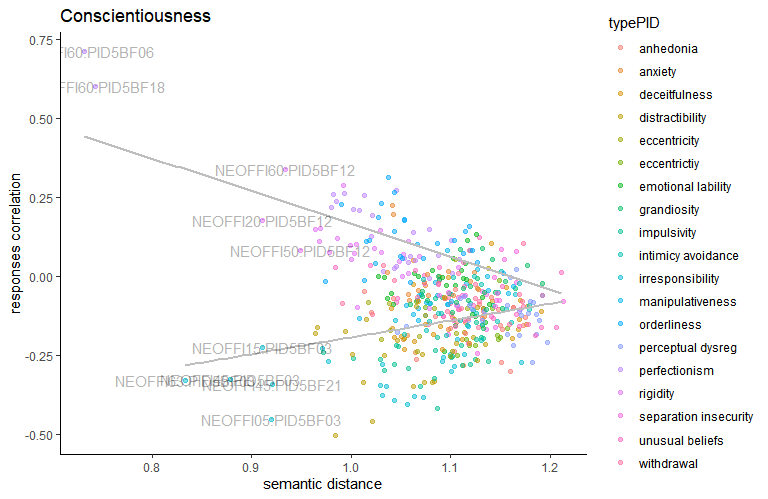
To obtain insight on the items driving these associations, we turn to examining each NEO subscale in turn, avoiding the excessive clutter from the large number of item pairs.

For conscientiusness, we can show that with lower semantic distance the associations diverge into two groups.



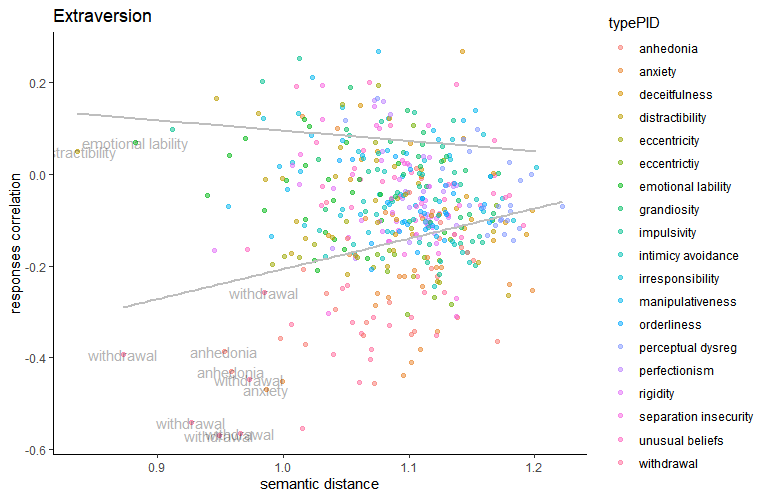
The items whose PID semantics drove a positive association with NEO conscientiousness items were those concerning rigidity and perfectionism, which are represented in two PID facets. In contrast, the semantics of the PID facets distractibility and irresponsibility drove the negative association with conscientiousness, meaning individuals who are more conscienscious are less distractable and irresponsible.

We now identify the items in the scale that are responsible for the negative association:

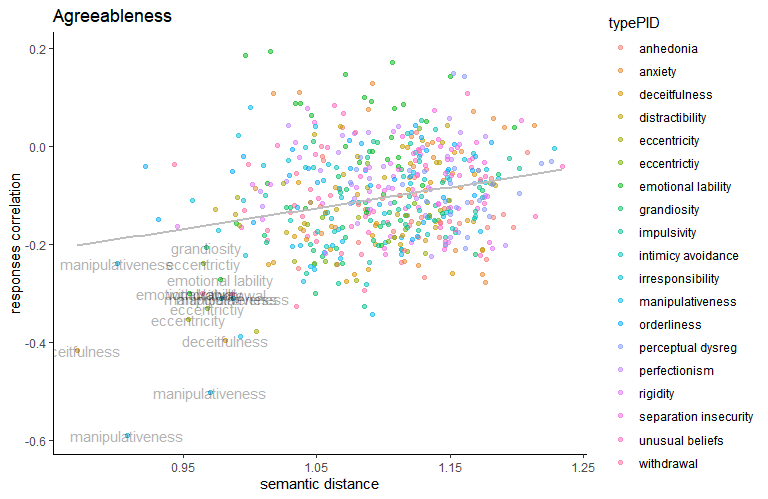


NEO item 60 is “I strive for excellence in everything I do” (Bei allem, was ich tue, strebe ich nach Perfektion”). This ends up correlated with responses on “I keep approaching things the same way, even when it isn’t working” (“Auch wenn es andere zum Wahnsinn treibt, bestehe ich darauf, alles perfekt zu machen”) and “I get stuck on one way of doing things, even when it’s clear it won’t work” (“Ich versuche Dinge weiter zu perfektionieren, auch wenn ich sie wahrscheinlich schon so gut wie möglich hinbekommen habe”). The differences in meaning between the German and English version are enough to produce different results in the semantic analysis: here, the notion of perfectionism drives a semantic association that is absent in the original English language (see Ontology.ipynb).

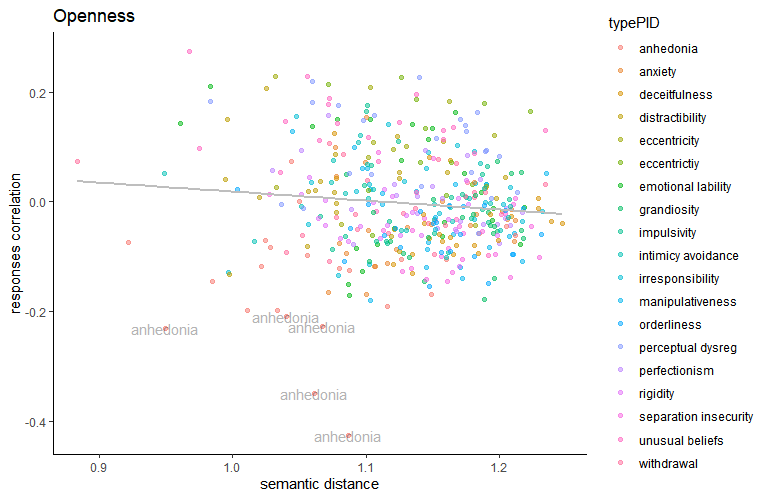
Extraversion, by contrast, appears to be so sparsely distributed because of the selective negative association of some items with anhedonia, anxiety, and withdrawal. Here, the language model detected semantic affinities between PID items concerning distractibility and emotional lability that were not confirmed by higher response associations.



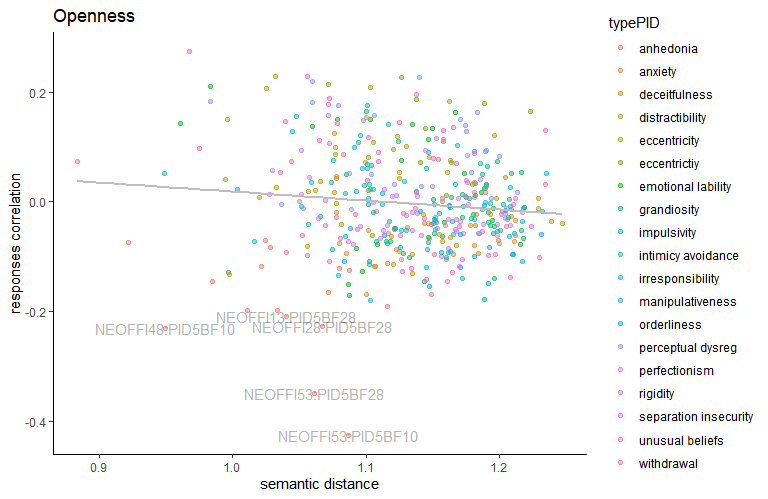
Agreableness, in contrast, presents a weak and less specific negative association with the PID traits, with some stronger associations with manipulativeness and deceitfulness.



Finally, we show what appears to be a genuine association between low openness and anhedonia. “Genuine”, because the language model detects no semantic affinity between the items of openness and anhedonia, but in some cases a negative association is present among participants.



We can identify the NEO items that are responsible for these associations:



NEOFFI53, which shows the most negative correlations with PID28 and PID10, is the text I have a lot of intellectual curiosity” (“Ich bin sehr wissbegierig”). The correlated anhedonia PID items are “I rarely get enthusiastic about anything” (“Ich bin selten von irgendetwas begeistert”) and “Nothing seems to interest me very much” (“Nichts scheint mich wirklich zu interessieren”). Here, the notion of intellectual curiosity appears to have been encoded as being semantically distinct from general enthusiasm and interest by the language model, but ends up correlating with anhedonia in the responses of participants.