

# Personality Profiling of Fictional Characters using Sense-Level Links between Lexical Resources

Lucie Flekova<sup>†</sup> and Iryna Gurevych<sup>†‡</sup>

<sup>†</sup> Ubiquitous Knowledge Processing Lab (UKP-TUDA)

Department of Computer Science, Technische Universität Darmstadt

<sup>‡</sup> Ubiquitous Knowledge Processing Lab (UKP-DIPF)

German Institute for Educational Research

[www.ukp.tu-darmstadt.de](http://www.ukp.tu-darmstadt.de)

*“Always be yourself, unless you can be Batman. Then always be Batman.”*

– Bill Murray

## Abstract

This study focuses on personality prediction of protagonists in novels based on the Five-Factor Model of personality. We present and publish a novel collaboratively built dataset of fictional character personality and design our task as a text classification problem. We incorporate a range of semantic features, including WordNet and VerbNet sense-level information and word vector representations. We evaluate three machine learning models based on the speech, actions and predicatives of the main characters, and show that especially the lexical-semantic features significantly outperform the baselines. The most predictive features correspond to reported findings in personality psychology.

## 1 Introduction

Recent progress in NLP has given rise to the field of personality profiling - automated classification of personality traits based on written, verbal and multimodal behavior of an individual. This research builds upon findings from classical personality psychology and has applications in a wide range of areas from medicine (suicide prevention) across security (forensics, paedophile detection, cyberbullying) to marketing and sales (recommendation systems, target group profiles). The gold standard labels for an objective evaluation of personality are mostly obtained by means of personality tests of the Five Factor Model (FFM) (McCrae and Costa, 1987; Goldberg, 1990), which is well-known and widely accepted in psychology and other research fields. The FFM defines personality

along five bipolar scales: Extraversion (sociable vs. reserved), Emotional stability (secure vs. neurotic), Agreeableness (friendly vs. unsympathetic), Conscientiousness (organized vs. careless) and Openness to experience (insightful vs. unimaginative). Psychologists have shown that these five personality traits are stable across individual lifespan, demographical and cultural differences (John and Srivastava, 1999) and affect many life aspects. (Terracciano et al., 2008; Rentfrow et al., 2011).

It has been shown that the personality traits of readers impact their literature preferences (Tirre and Dixit, 1995; Mar et al., 2009). Psychology researchers also found that perceived similarity is predictive of interpersonal attraction (Montoya et al., 2008; Byrne, 1961; Chartrand and Bargh, 1999). More explicitly, recent research (Kaufman and Libby, 2012) shows that readers of a narrative develop more favorable attitudes and less stereotype application towards a character, if his difference (e.g. racial) is revealed only later in the story. We therefore hypothesize that readers might have a preference for reading novels depicting fictional characters that are similar to themselves. Finding a direct link between reader’s and protagonist’s personality traits would advance the development of content-based recommendation systems. As a first step to explore this hypothesis further, it needs to be determined if we are able to construct a personality profile of a fictional character in a similar way as it is done for humans, and which aspects of personality profiling can be exploited to automatize such procedure.

In this paper, we open this research topic by presenting a novel collaboratively built dataset of fictional character personality in Section 3, which we make available on our website.<sup>1</sup> Framing the personality prediction as a text classification task, we incorporate features of both lexical-

<sup>1</sup><https://www.ukp.tu-darmstadt.de/data/personality-profiling/>

resource-based and vector space semantics, including WordNet and VerbNet sense-level information and vectorial word representations. We evaluate three machine learning models based on the speech (Section 4), actions (Section 5) and predicatives (Section 6) of the protagonists, and show that especially on the direct speech and action data the lexical-semantic features significantly outperform the baselines. Qualitative analysis reveals that the most predictive features correspond to reported findings in psychology and NLP.

## 2 Related work

Research in the the area of content-based recommendation systems have shown that incorporating semantic information is valuable for the user and leads to measurable improvements (Passant, 2010; Di Noia et al., 2012; Heitmann and Hayes, 2010). De Clercq et al. (2014) incorporated semantic frames from FrameNet into the recommendation system for books. They represent the plot of each book with a sequence of ca. 200 semantic frames and has shown that the frame information (such as Killing - Revenge - Death) outperforms the bag-of-words approach. Recent NLP experiments begin to reveal the importance of entity-centric models in a variety of tasks. Chambers (2013) show improvement in event schema induction by learning entity-centric rules (e.g., a victim is likely to be a person). Bamman et al. (2014) and Smith et al. (2013) present latent variable models for unsupervised learning of latent character types in movie plot summaries and in English novels, taking authorial style into account. However, even the state-of-the-art NLP work rather describes personas of fictional characters by their role in the story - e.g., action hero, valley girl, best friend, villain etc. - or by their relations to other characters, such as mother or daughter (Elson et al., 2010; Kokkinakis and Malm, 2011), rather than by their inner preferences and motivations. It is important to note here that determining a personality of a character is a very different task from determining its role in the story. Psychological understanding of personality, in contrast to role attribution requires a certain detached objectivity - even outright villains may have traits considered desirable in real life. For example, the devil has in many tales a very high aspiration level, appearing highly conscientious and agreeable. We hypothesize that these deeper personality aspects are

those which drive reader's affiliation to the character, thus deserve to be examined closer.

Also literary scholars formulate ad hoc personality descriptions for their experiments, for example to test hypotheses from evolutionary psychology (Johnson et al., 2011) or examine fictional portrayals of physicists (Dotson, 2009). These descriptions are usually adjusted to the experiment focus (e.g. emotions, relationships, ambitions). As McCrae et al. () point out, *a standard set of personality traits, that encompass the full range of characteristics found in all characters in literature* (p.77), is needed for a better comparison.

Hence we base our present study primarily on the previous NLP research on personality prediction of human individuals. Correlations between lexical and stylistic aspects of text and the five FFM personality traits of the author have been found in numerous experiments, with extraversion receiving the most attention (Pennebaker and King, 1999; Dewaele and Furnham, 1999; Gill and Oberlander, 2002; Mehl et al., 2006; Aran and Gatica-Perez, 2013; Lepri et al., 2010). The LIWC lexicon (Pennebaker et al., 2001) established its position as a powerful mean of such analysis.

The first machine learning experiments in this area were conducted by Argamon et al. (2005), Oberlander and Nowson (2006) and Mairesse et al. (2007). Researchers predicted the five personality traits of the authors of stream-of-consciousness essays, blog posts and recorded conversation snippets. Given balanced data sets, Mairesse et al. (2007) report binary classification accuracy of 50-56% on extraversion in text and 47-57% in speech, using word ngrams, LIWC, MRC psycholinguistic database (Coltheart, 1981) and prosodic features. Additional improvement is reported when the extraversion was labeled by external judges rather than by self-testing. Extended studies on larger datasets achieve accuracies around 55% (Nowson, 2007; Estival et al., 2007). More recent work in this area focuses on the personality prediction in social networks (Kosinski et al., 2013; Kosinski et al., 2014) and multimodal personality prediction (Biel and Gatica-Perez, 2013; Aran and Gatica-Perez, 2013). These trends emphasized the correlation of network features and audiovisual features with extraversion, giving rise to the Workshop on Computational Personality Recognition (for an overview see (Celli et al., 2013; Celli et al., 2014).

### 3 Data set construction

Traditionally, the gold standard for this supervised classification task is obtained by the means of personality questionnaires, used for the Five-Factor Model, taken by each of the individuals assessed. This poses a challenge for fictional characters. However, strong correlations have been found between the self-reported and perceived personality traits (Mehl et al., 2006). Our gold standard benefits from the fact that readers enjoy discussing the personality of their favourite book character online. A popular layman instrument for personality classification is the Myers-Briggs Type Indicator (Myers et al., 1985), shortly MBTI, which sorts personal preferences into four opposite pairs, or dichotomies, such as Thinking vs. Feeling or Judging vs. Perceiving. While the MBTI validity has been questioned by the research community (Pittenger, 2005), the Extraversion scale is showing rather strong validity and correlation to similar trait in the Five-Factor Model (McCrae and Costa, 1989; MacDonald et al., 1994). Our study hence focuses on the Extraversion scale.

Our data was collected from the collaboratively constructed Personality Databank<sup>2</sup> where the readers can vote if a book character is, among other aspects, introverted or extraverted. While the readers used codes based on the MBTI typology, they did not apply the MBTI assessment strategies. There was no explicit annotation guideline and the interpretation was left to readers' intuition and knowledge.<sup>3</sup> This approach of gold standard collection has several obvious drawbacks. First, the question is posed as dichotomic, while in reality the extraversion is a normally distributed trait in human population (Goldberg, 1990). Second, users can view the vote of previous participants, which may influence their decision. While we address both of these issues in our ongoing data collection project based on the Five-Factor Model, we consider them acceptable for this study due to the exploratory character of our pilot research.

We have collected extraversion ratings for 298 book characters, of which 129 (43%) are rather extraverted and 166 (56%) rather introverted. Rated

<sup>2</sup><http://www.mbti-databank.com/>

<sup>3</sup>MBTI defines extraversion as "getting energy from active involvement in events, having a lot of different activities, enjoying being around people." In the NEO Five-Factor Inventory (Costa and McCrae, 1992), underlying facets of extraversion are warmth, gregariousness, assertiveness, activity, excitement seeking and positive emotion.

characters come from a wide range of novels that the online users are familiar with, often covering classical literature which is part of the high school syllabus, as well as the most popular modern fiction, such as the Harry Potter series, Twilight, Star Wars or A Game of Thrones. A sample of the most rated introverts and extraverts is given in table 1. The rating distribution in our data is strongly U-shaped. The percentage agreement of voters in our data is 84.9%, calculated as:

$$P = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^k \frac{n_{ij}(n_{ij} - 1)}{n(n - 1)}$$

where  $k = 2$  (introvert, extravert),  $N$  is the number of book characters and  $n$  the number of votes per character. Voters on the website were anonymous and cannot be uniquely identified for additional corrections. There is no correlation between the extraversion and the gender of the character.

Character	Book	E	I
Tyrion Lannister	Game of Thrones	52	1
Cersei Lannister	Game of Thrones	48	7
Joffrey Baratheon	Game of Thrones	41	1
Ron Weasley	Harry Potter series	37	4
Jamie Lannister	Game of Thrones	38	9
Draco Malfoy	Harry Potter series	33	4
Anakin Skywalker	Star Wars series	30	6
Robert Baratheon	Game of Thrones	28	2
Gimli	Lord of the Rings	19	2
Jar Jar Binks	Star Wars series	12	2
Harry Potter	Harry Potter series	1	71
Severus Snape	Harry Potter series	1	65
Gandalf	Lord of the Rings	1	59
Yoda	Star Wars series	0	58
Jon Snow	Game of Thrones	1	47
Albus Dumbledore	Harry Potter series	4	46
Ned Stark	Game of Thrones	0	41
Aragorn	Lord of the Rings	1	41
Frodo	Lord of the Rings	1	40
Bran Stark	Game of Thrones	1	36

Table 1: Extraverts (E) and introverts (I) with the highest number of user votes.

Our set of English e-books covered 220 of the characters from our gold standard. We have built three systems to assess the following:

1. **Direct speech:** Does the style and content of character's utterances predict his extraversion in a similar way as it was shown for living individuals?
2. **Actions:** Is the behavior, of which a character is an agent, predictive for extraversion?
3. **Predicatives and adverbs:** Are the explicit (*John was an **exhibitionist***) or implicit (*John shouted **abruptly***) descriptions of the character in the book predictive for extraversion?

In the next three sections we present the experimental settings and results for each of the systems.

## 4 Direct speech of fictional characters

The system for the direct speech resembles the most to the previous systems developed for author personality profiling, e.g. on stream of consciousness essays (Mairesse et al., 2007) or social media posts (Celli et al., 2013) and therefore provides the best opportunity for comparison between human individuals and fictional characters. On top of the comparison to previous research, we exploit the sense links between WordNet and VerbNet to extract additional features - an approach which is novel for this type of task.

### 4.1 Extraction and assignment of speech

We process the book text using freely available components of the DKPro framework (Gurevych et al., 2007). The most challenging task in building the direct speech data set is assigning to the direct speech utterance the correct speaker. We benefit from the epub format of the e-books which defines a paragraph structure in such a way, that only the indirect speech chunk immediately surrounding the direct speech can be considered:

```
<p> John turned to Harry.
  "Let's go," he said.</p>
```

Given the large amount of text available in the books we focus on precision and discard all utterances with no explicit speaker (i.e., 30-70% of the utterances, dependent on the book), as the performance of current systems on such utterance types is still fairly low (O'Keefe et al., 2012; He et al., 2013; Iosif and Mishra, 2014). Similarly, conventional coreference resolution systems did not perform well on this type of data and were therefore not used in the final setup. We adapt the Stanford Named Entity Recognizer (Finkel et al., 2005) to consider titles (Mr., Mrs., Sir...) as a part of the name and to treat the first person I as a named entity. However, identifying only the named entity PERSON in this way is not sufficient. On our evaluation sample consisting of *A Game of Thrones* and *Pride and Prejudice* books (the former annotated by us, the latter by He et al. (2013)), 20% of utterances with explicit named speaker were not recognized. Of those correctly identified as a Person in the adjacent indirect speech, 17% were not the speakers. Therefore we implemented a

custom heuristics (Algorithm 1), which additionally benefits from the WordNet semantic classes of verbs, enriching the speaker detection by grabbing the nouns . With this method we retrieve 89% of known speakers, of which 92% is assigned correctly. Retrieved names are grouped based on string overlap (e.g. *Ser Jaime* and *Jaime Lannister*), excluding the match on last name, and corrected for non-obvious groupings (such as *Margaret* and *Peggy*).

---

#### Algorithm 1 Assign speaker

---

```
1: nsubj ← subjects in adjacent indirect speech
2: if count(nsubj(i) = PERSON) = 1 then speaker ← nsubj
3: else if count(nsubj(i) = PERSON) ≥ 1 then speaker ← the nearest one to directSpeech
4: else if directSpeech preceded by VERB.COMMUNICATION then speaker ← the preceding noun(s)
5: else if directSpeech followed by VERB.COMMUNICATION then speaker ← the following noun(s)
6: else if directSpeech followed by gap & VERB.COMMUNICATION then speaker ← the noun(s) in gap
7: else if directSpeech preceded by gap & VERB.COMMUNICATION then speaker ← the noun(s) in gap
return speaker
```

---

Our experimental data consists of usable direct speech sets of 175 characters - 80 extraverts (E) and 95 introverts (I) - containing 289 274 words in 21 857 utterances (on average 111 utterances for E and 136 for I, as I are often central in books).<sup>4</sup>

### 4.2 Classification approach for direct speech

All speech utterances of one book character are represented as one instance in our system. We use the leave-one-out classification setup due to the relatively small dataset size, using the support vector machines (SVM-SMO) classifier, which performs well on comparable tasks (Celli et al., 2013). The classification is performed through the DKPro TC Framework (Daxenberger et al., 2014).

**Lexical features** As a bottom-up approach we use the 1000 most frequent word uni-, bi- and trigrams, 1000 dependency word pairs, 1000 character trigrams and 500 most frequent verbs, adverbs, adjectives and interjections as binary features.

**Semantic features** Since the top-down approach, i.e. not focusing on individual words, has

---

<sup>4</sup>The data set size is comparable to ongoing personality profiling challenges - see <http://pan.webis.de>

been found more suitable for the personality profiling task on smaller data sets (Celli et al., 2013), we aim on capturing additional phenomena on a higher level of abstraction. The main part of our features is extracted on sense level. We use the most frequent sense of WordNet (Miller, 1995) to annotate all verbs in the direct speech (a simple but well performing approach for books). We then label the disambiguated verbs with their semantic field given in WordNet (WordNet defines 14 semantic classes of verbs which group verbs by their semantic field) and we measure frequency and occurrence of each of these classes (e.g. cognition, communication, motion, perception)<sup>5</sup>. Additionally, we use the lexical-semantic resource UBY (Gurevych et al., 2012) to access the WordNet and VerbNet information, and to exploit the VerbNet sense-level links which connects WordNet senses with the corresponding 273 main VerbNet classes (Kipper-Schuler, 2005). These are more fine-grained (e.g. pay, conspire, neglect, discover) than the WordNet semantic fields. WordNet covered 90% and VerbNet 86% of all the verb occurrences.

On word level, we extract 81 additional features using the Linguistic Inquiry and Word Count (LIWC) tools (Pennebaker et al., 2001), which consists of lexicons related to psychological processes (cognitive, perceptual, social, biological, affective) and personal concerns (achievement, religion, death...) and other categories such as fillers, disfluencies or swear words<sup>6</sup>. Additionally, since emotion detection has been found predictive in previous personality work (Mohammad and Kiritchenko, 2013), we measure overall positive and negative sentiment expressed per character, using SentiWordNet (Esuli and Sebastiani, 2006) and NRC Emotion Lexicon (Mohammad and Turney, 2010) for the word lookup, inverting sentiment scores for negated dependency sub-tree given by the Stanford Parser.

**Stylistic features** Features of this group capture the syntactic and stylistic properties of the utterances of a character, disregarding the content. Starting from the surfacial properties, we measure the sentence, utterance and word length, including the proportion of words shorter than 4 or longer than 6 letters, frequency of each punctuation mark,

<sup>5</sup><https://wordnet.princeton.edu/manual/lexnames.5WN.html>

<sup>6</sup>For complete overview refer to [www.liwc.net](http://www.liwc.net)

and endings of each adjective as per Corney et al. (2002). On the syntax level we measure the frequency of each part of speech as well as the 500 most frequent part-of-speech bi-, tri- and quadri-grams, and the frequency of each dependency obtained from the Stanford Parser. We additionally capture the frequency of superlatives, comparatives and modal verbs, the proportion of verbs in present, past and future tense, and the formality of the language as per the part-of-speech-based formality coefficient (Heylighen and Dewaele, 2002), and measure the average depth of the parse trees.

**Word embeddings as features** Since vector space semantics has been beneficial for predicting author’s personality in previous work (Neuman and Cohen, 2014), we use a pre-trained word vector model created by the GloVe algorithm (Pennington et al., 2014) on English Wikipedia. GloVe employs a global log-bilinear regression model that combines the advantages of the global matrix factorization and local context window methods. We assign the resulting 300-dimensional vectors to the words in character’s direct speech, excluding stopwords, and calculate an average vector for each character. We calculate for each test character the cosine similarity to the mean vector of extravert, resp. introvert, in the training data, and to each character in the training set individually using the DL4J NLP package<sup>7</sup>. We consider both the final scalar outcome and the difference of each of the individual vector dimensions as features.

### 4.3 Classification results on direct speech

Table 2 shows the precision, recall,  $F_1$ -score and accuracy for extraversion and introversion as a weighted average of the two class values.

ID	Feature set	P	R	F	A
1	-(baseline)	.295	.543	.382	.543
2	Ngrams	.519	.514	.515	.514
3	LIWC	.555	.560	<b>.552</b>	.560
4	WordNet	.527	.548	.528	.548
5	VerbNet	.649	.617	<b>.572</b>	.617
6	Style	.560	.581	<b>.558</b>	.581
7	Sentiment	.524	.543	.419	.543
8	Vectors	.295	.543	.382	.543
9	All	.550	.632	<b>.588</b>	.632
Percentage human agreement:					.849

Table 2: Weighted precision (P), recall (R), F-score (F) and accuracy (A) for a direct speech system, in each line using only the given group of features. WordNet stands for WordNet semantic labels, VerbNet setup uses the WordNet-VerbNet links to retrieve VerbNet labels. Highlighted F-scores differ from the majority baseline significantly ( $p < 0.05$ ), using an approximate randomization test.

<sup>7</sup><http://deeplearning4j.org/>

<b>Introvert</b>		
Feat.group	Features	Merit
unigrams	<i>reason, trouble, strange, indeed</i>	.24-.19
bigrams	<i>this time, tell me, I hope</i>	.19-.16
LIWC	Negate, Discrepancy, Insight, Exclusion	.18-.13
WordNet	stative, creation, cognition	.15-.09
VerbNet	lodge, hunt, defend	.23-.19
Style	modal verbs, neg, sbar, articles	.19-.14
<b>Extravert</b>		
Feat.group	Features	Merit
ngrams	<i>we, hurry, fat, dirty</i>	.24-.19
LIWC	We, Inclusion, Pronoun, Body	.18-.09
WordNet	motion, contact, communication, body, perception, change	.14-.07
VerbNet	get, talk, substance emission	.18-.15
Style	pronoun We, whadjp, type-token ratio., interjections	.20-.14

Table 3: The most predictive features for each group for speaker’s extraversion and introversion. Correlation merit, as per the correlation feature selection in WEKA, evaluates Pearson’s correlation between the feature and the class

Similarly to previous research (Mairesse et al., 2007; Celli et al., 2013), the bottom-up word based approach is outperformed by top-down semantic approaches which employ a more abstract feature representation. As in previous work, LIWC features exhibit good performance. However, the highest performance is achieved employing the VerbNet verb classes with WordNet word-sense disambiguation. Also stylistic features contribute substantially to the classification despite the mixture of genres in our book corpus - especially frequencies of modal verbs and part-of-speech ratios were particularly informative. The most predictive features from each group are listed in Table 3 together with their correlation merit (Hall, 1999), and compared with previous work in Table 4.

Feature	I/E	Ref	Feature	I/E	Ref
Predictive also in our data:			No effect in our data:		
Pronoun ‘we’	-/+	[3]	Neg. emot.	+/-	[1]
Tentative, unsure	+/-	[1]	Pos. emot.	-/+	[1]
Exclusive	+/-	[1]	Self-ref.	-/+	[1]
Inclusive	-/+	[1]	Formality	+/-	[2]
Insight	+/-	[1]	Elaborated	+/-	[3]
Nouns, articles	+/-	[2]	Long sent.	+/-	[3]
Lexical richness	+/-	[2]	Social	-/+	[3]
Negations	+/-	[2]			
Body functions	-/+	[2]			
Interjections	-/+	[3]			
<b>Source ID</b>		<b>Author</b>			
[1]		Pennebaker and King (1999)			
[2]		Dewaele and Furnham (1999)			
[3]		Mairesse et al. (2007)			

Table 4: Comparison of our results to previously reported predictive features for speaker’s extraversion (E), resp. introversion (I). We list publications where these features were, to our knowledge, reported as novel.

In accordance with the experiments of Pennebaker and King (1999), we observe more frequent exclusions (e.g. *without, but*), hedging and negation expressed by introverts, and inclusion (e.g. *with, and*) by extraverts. Extraverts talk more in first person plural, use more back-channels and interjections, and talk more about aspects related to their body. Introverts show more rationalization through insight words and more factual speech using less pronouns.

Additionally, the semantic features in Table 3 confirm the broad psychological characteristics of both types in general, i.e., for introverts the rationalization, uncertainty and preference for individual or rather static activities, and for extraverts their spontaneity, talkativeness and preference for motion. Furthermore, we observe certain directness in extraverts’ speech - note the predictive words *fat* and *dirty* and frequent descriptions of body functions.

**Discussion** Exploiting the links between lexical-semantic resources (performing WordNet word-sense disambiguation and using VerbNet verb classes linked to the disambiguated senses) was particularly beneficial for this task. WordNet semantic fields for verbs alone are too coarse-grained to capture the nuances in direct speech, and experiments with fine-grained VerbNet classes without WSD resulted in noisy labels. We did not confirm the previously reported findings on emotional polarity - we observe that the genre of the books (e.g. love romance vs horror story) have blurred the subtle differences between individual characters, unfortunately the dataset size did not allow for genre distinctions. Furthermore, a perceived extravert in our case can be a pure villain (Draco Malfoy, Joffrey Baratheon...) as well as a friendly companion (Gimli, Ron Weasley...), while the evil extravert types are possibly rarer in the experiments on human writing, or are more likely to fit under the MBTI definition of extraversion than FFM facets. Another potential cause, based on the error analysis, is the different target of the same sentiment for extraverts and introverts. For example, the ngram “I fear” is highly predictive for an introvert in our data while extraverts would rather use formulations to imply that others should fear. Similarly to Nowson et al. (2005), we did not find any difference in the formality measure of Heylighen and Dewaele (2002). Neither we did in the complexity of sentences as per the parse tree depth

and sentence length. It is probable that these aspects were also impacted by our broad variety of author style (F. Dostoyevsky vs J. K. Rowling). Our basic vector-based features carried no useful information in our case, in contrast to the personality research of Neuman and Cohen (2014). We observed that the factual content of the stories contributed to the character similarity measure more than the subtle personality differences.

## 5 Actions of fictional characters

While psycholinguists and consequently NLP researchers analyzed the relation between speech, resp. writing, and personality of an individual, psychologists often evaluate extraversion through behavioral personality questionnaires (Costa and McCrae, 1992; Goldberg et al., 2006). We hypothesize that similar behavior shall be predictive for extraversion of fictional characters as perceived by the readers.

### 5.1 Action extraction

For our purpose we define actions as the subject, verb and context of a sentence, where the subject is a named entity Person and the context is either a direct object in relation *obj* to the verb or a first child of the adjacent verb phrase in a parse tree. After grouping the actions per character, the subject name is removed. For example, a sample of actions of the character Eddard Stark of Game of Thrones would be: *X paused a moment, X studied his face, X changed his mind, X unrolled the paper, X said* etc., visualized in Figure 1. We obtained 22 030 actions for 205 characters (102 E, 116 I), with on average 100 actions for E and 101 for I. Note that also actions for those characters who do not talk enough in the books (often first-person perspectives) could be used.

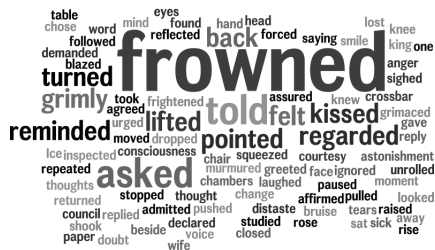


Figure 1: A revealing word cloud of the most frequent words from the actions of which Eddard Stark (Game of Thrones) is a subject. Size is proportional to the frequency of a word.

### 5.2 Action classification setup

In the system based on actions we use only a subset of the features described in 4.2. From the lexical features we focus on the 500 most frequent verbs and dependency word pairs. Semantic features are used the same way as in 4.2, profiting from LIWC, WordNet, Verbnet and the sentiment lexicons. Word embedding vectors for book characters are in this case computed by taking only the verbs into account rather than all content words. From the stylistic features we use the part-of-speech bigrams and trigrams, verb modality and verb tense.

### 5.3 Classification results on actions

Table 5 shows the performance of the classification models based on the protagonists' actions, using different feature groups. The overall performance is higher than for the direct speech model.

ID	Feature set	P	R	F	A
1	- (baseline)	.267	.517	.352	.517
2	Ngrams	.539	.506	.505	.507
3	LIWC	.600	.577	<b>.567</b>	.577
4	WordNet	.517	.518	.517	.518
5	VerbNet	.599	.583	<b>.578</b>	.583
6	Style	.573	.601	<b>.553</b>	.601
7	Sentiment	.357	.453	.382	.453
8	Vectors	.504	.497	.451	.497
9	All	.600	.623	<b>.598</b>	.623
Percentage human agreement:					.849

Table 5: Weighted precision (P), recall (R), F-score (F) and accuracy (A) for actions - in each line for a system using only the given group of features. WordNet stands for WordNet semantic labels, VerbNet setup uses the WordNet-VerbNet links. Highlighted F-scores differ from the majority baseline significantly ( $p < 0.05$ ), using an approx. randomization test.

Due to the lack of previous NLP experiments on this task, we compare our features to the actions measured in the International Personality Item Pool (Goldberg et al., 2006), frequently used personality assessment questionnaire (Table 6).

The most predictive features of this model capture the activity and excitement seeking facets of extraversion. Stylistic features reflect the complexity difference of the verb phrases (*John jumped vs. John thought about it*), extraverts being characterized by plain verbs. Semantic features exhibit higher precision than stylistic ones. Sense-linked semantic classes of VerbNet demonstrate the preference of extraverts for being active and expressing themselves - they *jump, fight, shout, run* in and run out, *eat and drink, see and hear* and get easily *bored*. Extraverts in books also

<b>Extravert</b> International Personality Item Pool: likes to party, feels comfortable around people, starts conversations, talks to many people, enjoys being a center of attention, makes friends easily, takes charge, captivates people, feels at ease with a company, is skilled in handling social situations Our experiment: bring (VN), consume (VN), contiguous location(VN), holding (VN), social (WN), motion (WN), emotion (WN) Leisure (LIWC), Home (LIWC), Family (LIWC), fight, march, care, take, jump, shriek, clear throat, bore, get to, come in, agree, hold, hear, inform, sell, come forward
<b>Introvert</b> International Personality Item Pool: Doesn't talk much, stays in the background, has little to say, does not draw attention, has difficulties to approach others, is quiet around strangers, feels uncomfortable around others, does not show feelings, is a private person, waits to be lead Our experiment: snooze (VN), conceal (VN), wish (VN), stative (WN), creation (WN), walk, sleep, lay, know, maintain, expect, hope, find out, might, help, explain

Table 6: Characteristic actions for extraverts and introverts as assessed in the IPIP personality questionnaire, compared to our most informative features

often *bring* or *hold* something. Introverts, on the other hand, seem to favor slow movements - while they are *thinking*, *reflecting*, *creating*, looking for *explanations* and *find out* solutions, they tend to *lie down*, *sit* or *walk*, eventually even *sleep* or *snooze*. The uncertainty typical for introverts is also notable in their actions, as they often *hope* or *wish* for something they *might* like to do. Additionally, semantic classes Social and Family, reported as correlated to extraversion by Pennebaker and King (1999) and not confirmed in our first model, became predictive in protagonists' actions.

## 5.4 Discussion

Also in this task, the VerbNet classes brought significant improvement in performance. The classification model based on actions outperforms not only the direct speech model, but also the state-of-the-art systems predicting authors' extraversion from the stream-of-consciousness essays (Mairesse et al., 2007; Celli et al., 2013; Neuman and Cohen, 2014). While surely not directly comparable, this result hints to the fact that the personality is easier to detect from behavior than from person's verbal expression. This would correspond to the findings of Mairesse et al. (2007), Biel and Gatica-Perez (2013) and Aran and Gatica-Perez (2013) on multimodal data sets.

## 6 Predicatives of fictional characters

Our third extraversion prediction system is subordinate to how fictional characters are described and to the manners in which they behave. We are not aware of a previous NLP work predicting extraversion using descriptive adjectives of the persons in question. We thus juxtapose the most predictive features of our system to the adjectival extraversion markers developed by Goldberg (1992).

### 6.1 Extraction of descriptive properties

In this setup we extract predicatives of the named entities PERSON in the books - relations *amod* (*angry John*) and *cop* (*John was smart*). As these explicit statements are very sparse in modern novels, we additionally include adverbial modifiers (*advmod*) related to person's actions (*John said angrily*). We extract data for 205 characters, with on average 43 words per character.

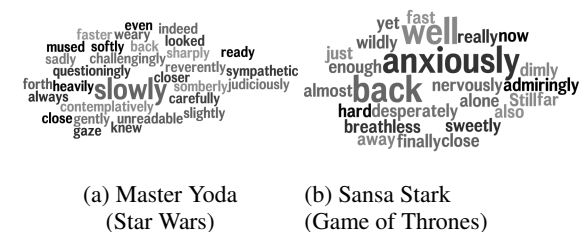


Figure 2: Frequency word clouds for character descriptions

### 6.2 Classification setup

This system uses similar set of lexical, semantic and vectorial features similarly as in 5.2, this time with the focus on adjectives, nouns and adverbs instead of verbs. Stylistic and VerbNet features are hence not included, word vectors are as in 4.2.

### 6.3 Classification results on descriptions

Table 7 reports on the performance of individual feature groups. With only few words per character semantic lexicons are less powerful than ngrams.

ID	Feature set	P	R	F	A
1	- (baseline)	.267	.517	.352	.517
2	Ngrams	.686	.657	<b>.648</b>	.657
3	LIWC	.645	.601	<b>.586</b>	.601
4	WordNet	.518	.545	.528	.545
5	Sentiment	.375	.463	.384	.463
6	Vectors	.267	.517	.352	.517
7	All	.692	.698	<b>.693</b>	.698
Percentage human agreement:					.849

Table 7: Weighted precision, recall, F-score and accuracy. Highlighted F-scores differ from the majority baseline significantly ( $p < 0.05$ ).



Table 8 displays the most predictive features in our system contrasted to the adjectival markers.

<b>Extravert</b>
Goldberg (1992) :
adventurous, mischievous, playful, rambunctious, dominant, forceful, demonstrative, exhibitionistic, flamboyant, brave, courageous, daring, assured,...
Our experiment :
excited, restlessly, stubbornly, restless, beloved, eager, abruptly, defiantly, darkly, eagerly, loudly, reluctant, stubborn, unwise, ruthless, quickly, abruptly, right, change (WN), social (WN)
<b>Introvert</b>
Goldberg (1992) :
bashful, shy, timid, inhibited, restrained unadventurous, unaggressive, uncompetitive bitter, joyless, melancholic, moody, morose,...
Our experiment :
anxious, patiently, hesitantly, backward, softly, warily, coldly, helplessly, respectfully, slowly, politely, thoughtfully, nervously, silent, carefully, gratefully, dryly, sheepishly, politely, weary, calm, gently, sadly, sideways, stative (WN)

Table 8: Characteristic adjectives for extraverts and introverts as reported by L. Goldberg, compared to our most informative features as per the correlation merit

#### 6.4 Discussion on errors

All our systems had issues with characters rated by less than five readers and with protagonists with low agreement. Other challenges arise from authorial style, age of the novel and speech individuality of characters (e.g. *Yoda*). Varied length of information for different characters poses issues in measuring normally distributed features (e.g. ratio of jumping verbs), being in shorter texts less reliable. Ongoing and future work on this task addresses the limitations of these initial experiments, especially the data set size and the gold standard quality. Extending the data will also enable us to examine different book genres as variables for the personality distribution and feature impact. It will be worth examining the relations between characters, since we observed certain patterns in our data, such as the main introvert character supported by his best friend extravert. Additionally, we want to verify if the system in Section 6 is overly optimistic due to the data size.

#### 7 Conclusion and future work

Automated personality profiling of fictional characters, based on rigorous models from personality psychology, has a potential to impact numerous domains. We framed it as a text classification problem and presented a novel collaboratively built dataset of fictional personality. We incor-

porate features of both lexical resource-based and vectorial semantics, including WordNet and VerbNet sense-level information and vectorial word representations. In models based on the speech and actions of the protagonists, we demonstrated that the sense-linked lexical-semantic features significantly outperform the baselines. The most predictive features correspond to the reported findings in personality psychology and NLP experiments on human personality. Our systems based on actions and appearance of characters demonstrate higher performance than systems based on direct speech, which is in accordance with recent research on personality in social networks (Kosinski et al., 2014; Biel and Gatica-Perez, 2013), revealing the importance of the metadata. We have shown that exploiting the links between lexical resources to leverage more accurate semantic information can be beneficial for this type of tasks, oriented to actions performed by the entity. However, the human annotator agreement in our task stays high above the performance achieved. Considering that most of the successful novels were produced as movies, we cannot exclude that our annotators based their decision on the multimodal representation of the protagonists. In the future we aim on collecting a more detail and rigorous gold standard through gamification and expanding our work on all five personality traits from the Five-Factor Model and their facets, and ultimately extend our system to a semi-supervised model dealing with notably larger amount of data. We also plan to examine closer the differences between perceived human and fictional personality, and the relationship between the personality of the reader and the characters.

#### Acknowledgments

This work has been supported by the Volkswagen Foundation as part of the Lichtenberg Professorship Program under grant No. I/82806 and by the German Research Foundation under grant No. GU 798/14-1. Additional support was provided by the German Federal Ministry of Education and Research (BMBF) as a part of the Software Campus program under the promotional reference 01-S12054 and by the German Institute for Educational Research (DIPF). We also warmly thank Holtzbrinck Digital GmbH for providing a substantial part of the e-book resources, and the EMNLP reviewers for their helpful comments.

## References

- Oya Aran and Daniel Gatica-Perez. 2013. Cross-domain personality prediction: from video blogs to small group meetings. In *Proceedings of the 15th ACM on International conference on multimodal interaction*.
- Shlomo Argamon, Sushant Dhawle, Moshe Koppel, and James W. Pennebaker. 2005. Lexical predictors of personality type. In *Proceedings of the Joint Annual Meeting of the Interface and the Classification Society of North America*.
- David Bamman, Ted Underwood, and Noah A. Smith. 2014. A bayesian mixed effects model of literary character. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*.
- Joan-Isaac Biel and Daniel Gatica-Perez. 2013. The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *Multimedia, IEEE Transactions on*, 15(1):41–55.
- Donn Byrne. 1961. Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology*, 62(3):713.
- Fabio Celli, Fabio Pianesi, David Stillwell, and Michal Kosinski. 2013. Workshop on computational personality recognition (shared task). In *Proceedings of the Workshop on Computational Personality Recognition*.
- Fabio Celli, Bruno Lepri, Joan-Isaac Biel, Daniel Gatica-Perez, Giuseppe Riccardi, and Fabio Pianesi. 2014. The workshop on computational personality recognition 2014. In *Proceedings of the ACM International Conference on Multimedia*. ACM.
- Nathanael Chambers. 2013. Event schema induction with a probabilistic entity-driven model. In *EMNLP*, volume 13, pages 1797–1807.
- Tanya L. Chartrand and John A. Bargh. 1999. The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893.
- Max Coltheart. 1981. The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, 33(4):497–505.
- Malcolm Corney, Olivier de Vel, Alison Anderson, and George Mohay. 2002. Gender-preferential text mining of e-mail discourse. In *Proceedings of 18th Annual Computer Security Applications Conference*. IEEE.
- Paul T. Costa and Robert R. McCrae. 1992. Professional manual: revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI). *Odessa, FL: Psychological Assessment Resources*.
- Johannes Daxenberger, Oliver Ferschke, Iryna Gurevych, and Torsten Zesch. 2014. Dkpro tc: A java-based framework for supervised learning experiments on textual data. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. System Demonstrations*, pages 61–66.
- Orphée De Clercq, Michael Schuhmacher, Simone Paolo Ponzetto, and Veronique Hoste. 2014. Exploiting framenet for content-based book recommendation. In *CBRecSys at ACM RecSys*, number 1613-0073, pages 14–21. CEUR-WS.
- Jean-Marc Dewaele and Adrian Furnham. 1999. Extraversion: The unloved variable in applied linguistic research. *Language Learning*, 49(3):509–544.
- Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, Davide Romito, and Markus Zanker. 2012. Linked open data to support content-based recommender systems. In *Proceedings of the 8th International Conference on Semantic Systems, ISEMANTICS*, pages 1–8.
- Daniel Dotson. 2009. Portrayal of physicists in fictional works. *CLCWeb: Comparative Literature and Culture*, 11(2):5.
- David K. Elson, Nicholas Dames, and Kathleen R. McKeown. 2010. Extracting social networks from literary fiction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*.
- Dominique Estival, Tanja Gaustad, Son Bao Pham, Will Radford, and Ben Hutchinson. 2007. Author profiling for english emails. In *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*.
- Andrea Esuli and Fabrizio Sebastiani. 2006. SentiWordNet: a publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6.
- Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 363–370. Association for Computational Linguistics.
- Alastair J. Gill and Jon Oberlander. 2002. Taking care of the linguistic features of extraversion. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*.
- Lewis R. Goldberg, John A. Johnson, Herbert W. Eber, Robert Hogan, Michael C. Ashton, C. Robert Cloninger, and Harrison G. Gough. 2006. The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, 40(1):84–96.

- Lewis R Goldberg. 1990. An alternative description of personality: the Big-Five factor structure. *Journal of personality and social psychology*, 59(6):1216.
- Lewis R Goldberg. 1992. The development of markers for the Big-Five factor structure. *Psychological assessment*, 4(1):26.
- Iryna Gurevych, Max Mühlhäuser, Christof Müller, Jürgen Steimle, Markus Weimer, and Torsten Zesch. 2007. Darmstadt Knowledge Processing Repository Based on UIMA. In *Proceedings of the First Workshop on Unstructured Information Management Architecture at Biannual Conference of the Society for Computational Linguistics and Language Technology*, Tübingen, Germany.
- Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 2012. Uby: A large-scale unified lexical-semantic resource based on LMF. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics.
- Mark A. Hall. 1999. *Correlation-based feature selection for machine learning*. Ph.D. thesis, The University of Waikato.
- Hua He, Denilson Barbosa, and Grzegorz Kondrak. 2013. Identification of speakers in novels. In *Proceedings of the 51st Annual Meeting on Association for Computational Linguistics*, pages 1312–1320.
- Benjamin Heitmann and Conor Hayes. 2010. Using linked data to build open, collaborative recommender systems. In *AAAI symposium Linked data meets artificial intelligence*.
- Francis Heylighen and Jean-Marc Dewaele. 2002. Variation in the Contextuality of Language: An Empirical Measure. *Foundations of Science*, 7(3):293–340.
- Elias Iosif and Taniya Mishra. 2014. From speaker identification to affective analysis: A multi-step system for analyzing children stories. *EACL 2014*, pages 40–49.
- Oliver P. John and Sanjay Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138.
- John A. Johnson, Joseph Carroll, Jonathan Gottschall, and Daniel Kruger. 2011. Portrayal of personality in victorian novels reflects modern research findings but amplifies the significance of agreeableness. *Journal of Research in Personality*, 45(1):50–58.
- Geoff F. Kaufman and Lisa K. Libby. 2012. Changing beliefs and behavior through experience-taking. *Journal of personality and social psychology*, 103(1):1.
- Karin Kipper-Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.
- Dimitrios Kokkinakis and Mats Malm. 2011. Character profiling in 19th century fiction. *Language Technologies for Digital Humanities and Cultural Heritage*.
- Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805.
- Michal Kosinski, Yoram Bachrach, Pushmeet Kohli, David Stillwell, and Thore Graepel. 2014. Manifestations of user personality in website choice and behaviour on online social networks. *Machine learning*, 95(3):357–380.
- Bruno Lepri, Ramanathan Subramanian, Kyriaki Kalimeri, Jacopo Staiano, Fabio Pianesi, and Nicu Sebe. 2010. Employing social gaze and speaking activity for automatic determination of the extraversion trait. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10, New York, NY, USA. ACM.
- Douglas A. MacDonald, Peter E. Anderson, Catherine I. Tsagarakis, and Cornelius J. Holland. 1994. Examination of the relationship between the Myers-Briggs Type Indicator and the NEO personality inventory. *Psychological Reports*, 74(1):339–344.
- François Mairesse, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research*.
- Raymond A. Mar, Keith Oatley, and Jordan B. Peterson. 2009. Exploring the link between reading fiction and empathy: Ruling out individual differences and examining outcomes. *Communications Journal*, 34(4):407–428.
- Robert R. McCrae and Paul T. Costa. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology*, 52(1):81.
- Robert R. McCrae and Paul T. Costa. 1989. Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality. *Journal of personality*, 57(1):17–40.
- Robert R. McCrae, James F. Gaines, and Marie A. Wellington. The five-factor model in fact and fiction. *Handbook of Psychology*.
- Matthias R. Mehl, Samuel D. Gosling, and James W. Pennebaker. 2006. Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology*, 90(5):862.

- George A. Miller. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.
- Saif M. Mohammad and Svetlana Kiritchenko. 2013. Using nuances of emotion to identify personality. *AAAI Technical Report WS-13-01 Computational Personality Recognition (Shared Task)*.
- Saif M. Mohammad and Peter D. Turney. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 26–34.
- R. Matthew Montoya, Robert S. Horton, and Jeffrey Kirchner. 2008. Is actual similarity necessary for attraction? a meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, 25(6):889–922.
- Isabel Briggs Myers, Mary H. McCaulley, and Robert Most. 1985. *Manual, a guide to the development and use of the Myers-Briggs type indicator*. Consulting Psychologists Press.
- Yair Neuman and Yochai Cohen. 2014. A vectorial semantics approach to personality assessment. *Scientific reports*, 4.
- Scott Nowson, Jon Oberlander, and Alastair J. Gill. 2005. Weblogs, genres and individual differences. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, volume 1666, page 1671. Citeseer.
- Scott Nowson. 2007. Identifying more bloggers: Towards large scale personality classification of personal weblogs. In *Proceedings of the International Conference on Weblogs and Social*.
- Jon Oberlander and Scott Nowson. 2006. Whose thumb is it anyway? classifying author personality from weblog text. In *Proceedings of the COLING/ACL on Main conference poster sessions*.
- Tim O’Keefe, Silvia Pareti, James R. Curran, Irena Koprinska, and Matthew Honnibal. 2012. A sequence labelling approach to quote attribution. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 790–799. Association for Computational Linguistics.
- Alexandre Passant. 2010. dbrec music recommendations using DBpedia. In *The Semantic Web–ISWC 2010*, pages 209–224.
- James W. Pennebaker and Laura A. King. 1999. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296.
- James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. *Proceedings of the Empirical Methods in Natural Language Processing (EMNLP 2014)*, 12.
- David J. Pittenger. 2005. Cautionary comments regarding the myers-briggs type indicator. *Consulting Psychology Journal: Practice and Research*, 57(3):210.
- Peter J. Rentfrow, Lewis R. Goldberg, and Daniel J. Levitin. 2011. The structure of musical preferences: a five-factor model. *Journal of personality and social psychology*, 100(6):1139.
- Noah A. Smith, David Bamman, and Brendan O’Connor. 2013. Learning latent personas of film characters. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*.
- Antonio Terracciano, Corinna E. Löckenhoff, Rosa M. Crum, O. Joseph Bienvvenu, and Paul T. Costa. 2008. Five-factor model personality profiles of drug users. *BMC Psychiatry*, 8(1):22.
- William C. Tirre and Sharvari Dixit. 1995. Reading interests: Their dimensionality and correlation with personality and cognitive factors. *Personality and Individual Differences*, 18(6):731–738.