The Internal Representations Questionnaire: Measuring modes of thinking



Hettie Roebuck¹ · Gary Lupyan¹

© The Psychonomic Society, Inc. 2020

Abstract

Does the format in which we experience our moment-to-moment thoughts vary from person to person? Many people claim that their thinking takes place in an inner voice and that using language outside of interpersonal communication is a regular experience for them. Other people disagree. We present a novel measure, the Internal Representation Questionnaire (IRQ) designed to assess people's subjective mode of internal representations, and to quantify individual differences in "modes of thinking" along multiple factors in a single questionnaire. Exploratory factor analysis identified four factors: Internal Verbalization, Visual Imagery, Orthographic Imagery, and Representational Manipulation. All four factors were positively correlated with one another, but accounted for unique predictions. We describe the properties of the IRQ and report a test of its ability to predict patterns of interference in a speeded word-picture verification task. Taken together, the results suggest that self-reported differences in how people internally represent their thoughts relates to differences in processing familiar images and written words.

Keywords Internal representations \cdot Inner voice \cdot Verbal representation \cdot Cognitive style \cdot Learning preference \cdot Thought \cdot Language

Just as perception is often accompanied by phenomenology it feels like something to imagine seeing a bicycle or to imagine hearing Für Elise—our internal thoughts often have a phenomenological character. What does it feel like to think through a problem? To recall last week's party? To imagine a friend's face? Most of us can say something about what these internal processes feel like to us. It is tempting to assume that this phenomenology is universal. However, evidence to date suggests that this is not the case. In 1880, Galton published the results of a survey of individual differences in visual imagery, asking 100 British men (including 19 Fellows of the Royal Society) to comment on their ability to visually imagine various kinds of information. Responses ranged from a very developed ability: "Thinking of the breakfast table this morning, all the objects in my mental picture are as bright as the actual

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13428-020-01354-y) contains supplementary material, which is available to authorized users.

Hettie Roebuck hroebuck@wisc.edu scene" to none: "My powers are zero. To my consciousness there is almost no association of memory with objective visual impressions. I recollect the breakfast table, but do not see it." (Galton, 1880, pp 304-306)¹

Contemporary work on individual differences in visual imagery has continued to use survey instruments similar to those used by Galton, e.g., the Vividness of Visual Imagery Questionnaire; VVIQ (Marks, 1973) and the Object-Spatial Imagery Questionnaire; OSIQ (Blajenkova, Kozhevnikov, & Motes, 2006) confirming the existence of large and stable individual differences (Amedi, Malach, & Pascual-Leone, 2005; Hatakeyama, 1997; McKelvie, 1994; McKelvie & Demers, 1979; Zeman, Dewar, & Della Sala, 2015) and a growing understanding of their consequences for behavior and their neural bases (Cui, Jeter, Yang, Montague, & Eagleman, 2007; Keogh & Pearson, 2018).

Here, we focus on another, much-less-studied aspect of phenomenology: the tendency to experience thoughts in the

¹ Department of Psychology, University of Wisconsin, Madison, WI, USA

¹ Galton concluded that "an over-readiness to perceive clear mental pictures is antagonistic to the acquirement of habits of highly generalized and abstract thought". The idea that imagery is antagonistic to abstraction has not been supported by later work (Brewer & Schommer-Aikins, 2006). One wonders whether Galton's own self-admitted lack of vivid imagery may have contributed to his conclusion.

form of language, i.e., internal verbalization. Many people describe that their thinking takes the form of an inner voice (Alderson-Day & Fernyhough, 2015a; Hurlburt, Heavey, & Kelsey, 2013; Klinger & Cox, 1987). As with visual imagery, earlier surveys have revealed large individual differences in the propensity to hear an "inner voice" (Heavey & Hurlburt, 2008; Heavey et al., 2018) as well as differences in the quality of the inner voice and situations in which it is experienced (Beyler & Schmeck, 1992; Kosslyn, Brunn, Cave, & Wallach, 1984; Macleod, Hunt, & Mathews, 1978). The range of people's experiences is captured by the following comments on a recent Reddit thread:

"I have a very strong inner voice that is constantly narrating my thoughts, what I write or the stuff that I read" (kafoBoto, 2016) "My thoughts don't 'sound' like anything. . . in fact I was pretty blown away when people said they literally had an inner voice—I thought it was a figure of speech" (Kunkletown, 2012).

Here, we present a new instrument—the internal representation questionnaire (IRQ)—for measuring individual differences in internal verbalization and other modes of thinking within a single questionnaire. We also set out to test the questionnaire's usefulness in predicting performance on an objective task that required participants to match meanings of objects expressed as words and images.

Why care about differences in internal verbalization?

Our interest in measuring people's internal verbalization stems from the idea that human cognition is augmented by language (Bowerman & Levinson, 2001; Lupyan, 2012b, 2016). Words—and language more broadly—have unique properties not present in other representational modalities making them a useful interface for manipulating mental states (Carruthers, 2002; Clark, 1997; Lupyan & Bergen, 2016). For example, language is categorical and compositional in a way that perception is not. The objects of our perception are always specific (e.g., a particular dog, a particular dinner, a particular hue). In contrast, the words "dog", "dinner", "green" are all categorical, abstracting away from perceptual details (to different degrees depending on the abstractness of the word in question). Mental representations constructed with the aid of, or augmented by, language may therefore take on a more categorical form facilitating inference and certain types of reasoning (Baldo, Paulraj, Curran, & Dronkers, 2015; Boutonnet & Lupyan, 2015; Holmes & Wolff, 2013;

Lupyan, 2015; Lupyan & Thompson-Schill, 2012). A more salient inner voice may reflect a greater involvement of language in the construction of mental representations and may help to account for different profiles that people show in categorization, memory, and reasoning tasks (e.g., Wasserman & Castro, 2012). Alternatively, these subjective differences may be epiphenomenal with no functional consequences.

How does the IRQ differ from existing instruments?

We are far from the first to attempt to quantify individual differences in inner verbalization or inner speech. Previously developed inner speech instruments include Duncan and Chevene's Self-Verbalization Questionnaire (1999), Brinthaupt et al.'s Self-Talk Scale (Brinthaupt, Hein, & Kramer, 2009), Calvete et al's Self-Talk Inventory (Calvete et al., 2005), Siegrist's Inner-Speech Scale (Siegrist, 1995), Fernyhough and colleagues' Varieties of Inner Speech Questionnaire (Alderson-Day, Mitrenga, Wilkinson, McCarthy-Jones, & Fernyhough, 2018; McCarthy-Jones & Fernyhough, 2011), and the Nevada Inner Speech Questionnaire (Heavey et al., 2018). These assessments tend to have relatively high test-retest reliability, but different instruments correlate poorly, suggesting low convergent validity (Uttl, Morin, & Hamper, 2011). The issue is that despite the similarity in the names of the tests (inner-verbalization, innerspeech, self-talk), what these tests assess varies dramatically from test to test. Many include statements relating to rumination and self-evaluation, e.g., "I hear my mother's voice criticizing me in my mind." from an Evaluative/Motivational factor (The Varieties of Inner Speech Questionnaire; McCarthy-Jones & Fernyhough, 2011), or "If I am not feeling well, I often talk to myself about my state." (The Inner Speech Scale; Siegrist, 1995). Some instruments were designed with the explicit goal of distinguishing different forms of innerspeech, e.g., "My thinking in words is more like a dialogue with myself, rather than my own thoughts in a monologue." (Alderson-Day et al., 2018). Other popular instruments such as the Verbal-Visualizer Questionnaire (VVQ; Kirby, Moore, & Schofield, 1988) derive "verbal" factors that conflate preferences with self-rated abilities, cf. "I enjoy learning new words," "I read rather slowly".

In constructing the IRQ, our principal goal was to measure people's *propensity* to use internalized language in different situations that do not involve communication with other people. These include using language as a retrieval cue for autobiographical memories, for cueing a visualization of a scene (e.g., visualizing a beach by internally using the word "beach"), and talking with oneself when trying to work out a problem. A second goal was to have a single instrument that assesses differences in verbalization together with previously described differences, e.g., in visual imagery. To this end, we have incorporated into the IRQ items from previously validated measurements of visual imagery and added other forms of imagery such as auditory and tactile. By including these items, it allows us to see how these different forms of imagery relate to one another. In constructing the IRQ, we deliberately focus on differences in propensities (or tendencies) of the modes of representation rather than differences in abilities (one exception is the inclusion of a representational-manipulation factor which includes questions that involve self-rated abilities, see below for more details). Although sometimes correlated, abilities and propensities are distinct constructs. For example, someone who reports that they rarely see vivid images in their mind's eye may, when called upon by a specific task, be able to form a highly vivid image. Someone who reports not experiencing an internal voice may nevertheless be highly accurate in performing tasks such as rhyme judgments that one might expect to benefit from an inner voice (Langland-Hassan, Faries, Richardson, & Dietz, 2015). The inclusion of items relating to internal verbalization and different forms of imagery in a single instrument means that we are not assuming a tradeoff between verbal and nonverbal (e.g., visual) approaches. This assumption underlies much of the literature on "learning styles" (e.g., Mayer & Massa, 2003; Jonassen & Grabowski, 1993; Riding & Cheema, 1991).²

Constructing the Internal Representations Questionnaire (IRQ)

In designing the IRQ, we followed standard guidelines for designing psychometric scales (Clark & Watson, 1995; Simms, 2008). We began with a substantive validity phase, developing an initial pool of items, exploring their factor analytic structure, and modifying the items to maximize convergent and discriminant validity. We measured internal validity by re-testing participants two months after an initial completion of the IRQ, and then carried out confirmatory factor analysis on another sample. As a measure of the IRQ's external validity, we correlated it with two published assessments of internal verbalization. Lastly, we tested the predictive validity of the IRQ using a speeded word–picture verification task. The goal of which was to determine if people with different IRQ profiles perform differently when required to match images to words and vice-versa.

Substantive validity phase

We began with a list of 81 statements about how people experience different kinds of thought processes. For example, "I hear words in my "mind's ear" when I think." and "If I am walking somewhere by myself, I often have a silent conversation with myself". Statements emphasized tendencies/propensities, although some statements about self-rated abilities were also included. To assess differences in visual imagery, the 81 initial items included statements from previously used visual imagery questionnaires including the (VVQ: Kirby et al., 1988), (VVIQ: Marks, 1973), and the Object-Spatial Imagery Questionnaire (OSIQ: Blajenkova et al., 2006). Each statement was presented together with a five-point Likert scale from Strongly Disagree to Strongly Agree. The original list of 81 statements is available at https://osf.io/ 8rdzh/ For this first phase, we recruited 180 students from the University of Wisconsin Madison. To help ensure compliance, we interspersed three attention-check questions among the 81 statements: "The word 'hotel' has three letters", "Elephants are larger than dogs" and "Most people have five legs". We excluded participants who incorrectly responded to any of these statements. Statement order was randomized.

We then conducted an exploratory factor analysis to measure the dimensionality of the resulting items. Question items that did not correlate higher than .30, or correlated above .90 with other items were excluded or rephrased. We constructed new statements to fill in any resulting gaps.

A refined 60-item scale was then administered to a sample of 222 adults recruited from Amazon's Mechanical Turk. Participants were aged between 20 and 72 (123 male, 96 female, three categorized as other/preferred not to state, mean age 36; SD 11 years). Oblique rotation was carried out for the analysis due to factor correlations (see Table 1). Four factors were retained in the model, based on the point of inflection in the scree plot. We included questions that had loadings greater than .40 on one factor and loadings less than .40 on all other factors. We also excluded any items if their exclusion increased Cronbach's alpha. We assessed homogeneity through inter-item correlations and included items only if they correlated with other items in their factor at greater than .30. The final set contained a total of 36 statements grouped into four factors.

The final set of questions and their loadings are shown in Table 2. The Visual Imagery factor consisted of ten items that described some aspect of visual/pictorial imagery (e.g., "I can close my eyes and easily picture a scene that I have experienced"). The Internal Verbalization factor included 12 items relating to experiencing thought in a spoken "inner voice" i.e., internally hearing words (e.g., "I think about problems in my mind in the form of a conversation with myself"). A—surprising—factor that emerged involved visualization of orthography. The Orthographic Imagery factor consisted of six items,

 $[\]overline{}^2$ The idea of a single verbalizer-visualizer dimension has led educational researchers to design curricula based on these two putatively opposed learning styles (Sternberg, Grigorenko, & Zhang, 2008). Not only is there no evidence that such curricula help learning (Pashler, McDaniel, Rohrer, & Bjork, 2008), but the very concept of a single visualizer-verbalizer dimension is routinely contradicted by *positive* relationships between assessments of verbalization and visualization (e.g., Leutner & Plass, 1998; Kraemer, Rosenberg, & Thompson-Schill, 2009), a relationship that we replicate here.

	Visual	Verbal	Orthographic	Cronbach's alpha	Test-retest reliability	Split- half
Visual	~	~	~	0.86	0.78	0.75
Verbal	.47**	~	~	0.86	0.68	0.79
Orthographic	.35**	.38**	~	0.72	0.65	0.48
Manipulation	.42**	.29**	.31**	0.79	0.64	0.71

Table 1. Correlations, Cronbach's alphas, test-rest reliability and split half correlations for each IRQ factor

Table 2. Items and their loadings on each IRQ factor

	Visual	Verbal	Orthographic	Manipulation
I often enjoy the use of mental pictures to reminisce	0.785	0.060	0.101	0.086
I can close my eyes and easily picture a scene that I have experienced	0.753	0.026	0.139	0.096
My mental images are very vivid and photographic	0.740	0.178	0.230	0.018
The old saying "A picture is worth a thousand words" is certainly true for me	0.685	0.095	0.066	0.206
When I think about someone I know well, I instantly see their face in my mind	0.680	0.023	0.120	0.063
I often use mental images or pictures to help me remember things	0.665	0.097	0.140	0.074
My memories are mainly visual in nature	0.638	0.046	0.001	0.004
When traveling to get to somewhere I tend to think more visually than verbally	0.612	0.153	0.010	0.001
If I talk to myself in my head it is usually accompanied by visual imagery	0.542	0.077	0.332	0.166
If I imagine my memories visually they are more often static than moving	0.418	0.043	0.137	0.220
I think about problems in my mind in the form of a conversation with myself	0.141	0.776	0.105	0.014
If I am walking somewhere by myself, I often have a silent conversation with myself	0.145	0.719	0.068	0.095
If I am walking somewhere by myself, I frequently think of conversations that I've recently had	0.141	0.699	0.132	0.035
My inner speech helps my imagination	0.156	0.653	0.078	0.103
I tend to think things through verbally when I am relaxing	0.024	0.650	0.132	0.121
When thinking about a social problem, I often talk it through in my head	0.135	0.595	0.005	0.043
I like to give myself some down time to talk through thoughts in my mind	0.298	0.583	0.036	0.342
I hear words in my "mind's ear" when I think	0.185	0.559	0.032	0.321
I rarely vocalize thoughts in my mind *	0.064	0.493	0.307	0.025
I often talk to myself internally while watching TV	0.062	0.486	0.325	0.103
My memories often involve conversations I've had	0.087	0.486	0.144	0.046
When I read, I tend to hear a voice in my "mind's ear"	0.050	0.467	0.001	0.180
When I hear someone talking, I see words written down in my mind	0.128	0.049	0.789	0.080
I see words in my "mind's eye" when I think	0.163	0.081	0.696	0.089
When I am introduced to someone for the first time, I imagine what their name would look like when written down	0.192	0.054	0.691	0.092
A strategy I use to help me remember written material is imagining what the writing looks like	0.071	0.057	0.618	0.039
I hear a running summary of everything I am doing in my head	0.156	0.326	0.616	0.012
I rehearse in my mind how someone might respond to a text message before I send it	0.038	0.363	0.400	0.132
I can easily imagine and mentally rotate three-dimensional geometric figures	0.107	0.226	0.064	0.670
I can easily choose to imagine this sentence in my mind pronounced unnaturally slowly	0.030	0.215	0.146	0.597
In school, I had no problems with geometry	0.252	0.136	0.216	0.594
It is easy for me to imagine the sensation of licking a brick	0.160	0.185	0.103	0.530
I find it difficult to imagine how a three-dimensional geometric figure would exactly look like when rotated *	0.158	0.302	0.014	0.526
I can easily imagine someone clearly talking, and then imagine the same voice with a heavy cold	0.080	0.051	0.183	0.502
I think I have a large vocabulary in my native language compared to others	0.143	0.181	0.145	0.476
I can easily imagine the sound of a trumpet getting louder	0.096	0.209	0.031	0.474

loading most on items that probed visualizing language as it is written, e.g., ("When I hear someone talking, I see words written down in my mind" and "I see words in my "mind's eye" when I think"). Some of the items we originally included to probe visuo-spatial imagery (known to be distinct from visual imagery of objects/faces) ended up clustering with non-visual items pertaining to manipulating mental representations more generally, e.g., responses to "I can easily imagine and mentally rotate three-dimensional geometric figures" were clustered with "It is easy for me to imagine the sensation of licking a brick" and "I can easily imagine the sound of a trumpet getting louder". We refer to this factor as Representational Manipulation. We term these factors "Visual Imagery" "Internal Verbalization", "Orthographic Imagery" and "Manipulational Representation", respectively. Note that Visual Imagery, Internal Verbalization, and Orthographic Imagery factors measure "propensities" while the items included in the Representational Manipulation factor tend to probe self-ratings of abilities.

The distributions of mean responses are shown in Fig. 1 (1 = lowest loading; 5 = highest loading; after correcting for reverse-coding). All factors were positively correlated with one another (Table 1). These correlations clearly contradict the still popular (but never empirically validated) idea that verbalization is somehow inversely related to visualization (e.g., Mayer & Massa, 2003).

To determine the internal consistency reliability, i.e., how well the items within a factor are related to each other, we measured Cronbach's alpha scores. Cronbach's alphas for each factor was >.70. See Table 3 for Cronbach's alphas of the final factors. We then assessed internal reliability of each factor using split-half analysis (Table 3). The split-half correlation for the IRQ overall was .71. The questionnaire was then retested on 125 of the original 222 participants, recruited through Mechanical Turk between 65 and 74 days later. Test–re-test reliability correlations are shown in Table 3.

External validity phase

Relationship to other scales To assess the validity of the new scale, 232 new participants completed the IRQ together with several other questionnaires designed to probe inner speech: The varieties of inner speech questionnaire (VISQ: McCarthy-Jones & Fernyhough, 2011), is designed to measure four main phenomenological properties of inner speech including dialogicality, condensed quality, evaluative/motivational nature, and the extent to which inner speech incorporates other people's voices. The self-talk scale (STS: Brinthaupt et al., 2009), is designed to measure self-talk across four factors: social assessment, reinforcement, criticism, and management. We also included the Need for Cognition scale (NFC: Cacioppo & Petty, 1982), designed to measure the extent to

which individuals are inclined towards effortful cognitive activities. The correlations between all the instruments are shown in Table 3. The IRQ's Internal Verbalization factor was positively correlated with other inner-speech instruments. The strongest correlation (.72) is with the dialogic factor of the VISQ which includes statements such as "My thinking in words is more like a dialogue with myself, rather than my own thoughts in a monologue". The IRQ's verbal factor was correlated to a lesser degree with VISQ's "other speech" factor (which includes questions specifically probing whether people's inner speech includes other people's voices) and had a small negative correlation (r = -.17) with VISQ's condensedspeech factor (e.g., "I think to myself in words using brief phrases and single words rather than full sentences"). Together this pattern suggests that the IRQ's Internal Verbalization factor is aligned more with people's propensity to engage themselves in an inner dialogue rather than experiencing an inner voice more generally. Intriguingly, the Internal Verbalization factor also showed an appreciable correlation (r = .67) with the Evaluative/Motivations factor of the VISQ which includes questions such as "I evaluate my behavior using my inner speech. For example, I say to myself, 'that was good' or 'that was stupid'", even though our questionnaire did not include any questions probing the use of inner speech for evaluative or motivational purposes. The Internal Verbalization factor was correlated with all four of the factors of the STS with strongest positive correlations with the Social (e.g., "I try to anticipate what someone will say and how I'll respond to him or her.") and Management factors (e.g., I'm giving myself instructions or directions about what I should do or say). The visual imagery factor was not correlated with any factor of the STS or VISQ. Intriguingly, the Orthographic factor was correlated with the "Other People" factor of the VISQ which probes people's experiences about hearing other people's speech in their mind's ear. In other work, we have found that our Orthographic factor is consistently correlated with some other measures that relate to heightened sensitivity to how others view oneself. For example, it is correlated with the "Reading for Recognition" dimension of the Dimensions of Reading Motivation questionnaire (Schutte & Malouff, 2007), which includes items such as "It is important to me to have others remark on how much I read." It is conceivable that our Orthographic factor is capturing some aspect of sensitivity to how the participant is perceived by others. This factor emerged from the factor analytic approach and has high internal consistency, but we hesitate to over-interpret its significance or the mechanisms that underlie it given its post hoc nature. Finally, the Manipulation factor showed a significant correlation with the Need for Cognition scale (e.g., "I find satisfaction in deliberating hard and for long hours"). Need for cognition is known to be positively correlated with general intelligence (Hill et al., 2013) and in our own work we have found that the Manipulation factor is positively correlated



Fig. 1 The distribution of loadings for individual scorers on each IRQ factor. Mean scores of 1 indicate strong disagreement; scores of 5 indicate strong agreement (after adjusting for reverse-coded items). To help visualize the individual IRQ profiles, we connected each participants' mean

high variation (as measured by the coefficient of variation across the four factors

responses with a line and color-coded participants with low, medium, and

with performance on Raven's matrices. It is possible that differences in the Manipulation Factor and Need for Cognition are caused by some more general intelligence factor. Alternatively, individuals with a greater capacity for/ proficiency with manipulating mental representations may be better at reasoning (hence the correlation with intelligence tests) and more likely to seek out information. Another possibility is that greater information-seeking behavior is the

Table 3. Correlations between the four IRQ factors with previously published measures of inner speech. The * denotes significance at p < .001

		STS					VISQ		NFC
	Social	Reinforcement	Criticism	Management	Condensed	Dialogic	Other People	Evaluative/ Motivational	Need for Cognition
Verbal	0.65*	0.42*	0.59*	0.67*	- 0.18	0.72*	0.35*	0.67*	0.14
Visual	0.04	0.16	0.01	0.17	- 0.03	-0.02	0.11	0.12	0.31*
Orthographic	0.3*	0.3*	0.25*	0.18	0.1	0.33*	0.51*	0.31*	0.03
Manipulation	0.16	0.06	0.14	0.22	- 0.09	0.13	0.16	0.1	0.42*
Social		0.51*	0.68*	0.8*	- 0.16	0.56*	0.2	0.56*	0.07
Reinforcement	0.51*		0.52*	0.57*	0.02	0.38*	0.23*	0.48*	0.08
Criticism	0.68*	0.52*		0.66*	-0.17	0.54*	0.21	0.57*	0.1
Management	0.8*	0.57*	0.66*		-0.17	0.56*	0.12	0.59*	0.16
Condensed	-0.16	0.02	-0.17	-0.17		- 0.2	0.24*	- 0.18	- 0.18
Dialogic	0.56*	0.38*	0.54*	0.56*	- 0.2		0.37*	0.64*	0.06
Other People	0.2	0.23*	0.21	0.12	0.24*	0.37*		0.36*	0.05
Evaluative/Motivational	0.56*	0.48*	0.57*	0.59*	- 0.18	0.64*	0.36*		0.06
Need for Cognition	0.07	0.08	0.1	0.16	-0.18	0.06	0.05	0.06	

primary causal factor. Distinguishing between these alternatives is beyond the scope of this paper.

Relationship to perceptions of others Do people attribute their modes of thinking to other people? We were curious if people who scored higher on the Internal Verbalization factor thought that others experience as much inner speech as they do. As part of one of the IRQ administrations, participants judged what percent of people experience their thoughts in the form of speech. The correlation between participants propensity to internally verbalize as estimated by the IRQ and their response to this question was r = .35, p < .001. Participants at the high end of verbalization (75th percentile) estimated that 65% of the population (SD = 29) experience inner speech. Participants with a low propensity to internally verbalize (25th percentile) estimated that only 39% do (SD = 29).

Confirmatory stage

We fit the confirmatory factor analysis model using lavaan (Rosseel, 2012) in R using new participants that were not included at the exploratory stage; 871 participants (232 participants on Mechanical Turk, and a further 639 from the student population). We used maximum likelihood estimation with full information maximum likelihood (FIML) for missing data. We standardized latent factors allowing free estimation of all factor loadings. The model had good fit: RMSEA of .052 90% CI (.049, .054). The full four-factor model fit the data significantly better than a single-factor solution (X 2)(6)=1825.3, p < .001), or a four-factor solution that did not allow covariance among the four latent factors $(X \ 2 \ (6) =$ 506.64, p < .001). The indicators all showed significant factor loadings. These results are consistent with the characterization of distinct factors for visual imagery, internal verbalization, orthographic imagery and representational manipulation.

Predictive validity: Does the IRQ predict objective behavior?

Having shown that the IRQ reveals substantial individual differences that are fairly stable within individuals across time, we next measured whether people's IRQ profiles predict performance on a more objective task. The task we chose was speeded word–picture verification. This task requires people to indicate whether a cue (for example, the written word "dog" or a picture of a dog) matches the target (an image of a dog, or the word "dog"). A picture of a dog followed by the word "cat" does not. By examining the speed with which people match targets to cues (for different types of cue-target relationships; see Fig. 2), we can measure whether a propensity to internally verbalize leads to a greater activation of phonology (e.g., Langland-Hassan et al., 2015; Kraemer et al., 2009). We can also examine how differences in the other IRQ factors relate to performance.

Predictions

We predicted that people who scored higher on the Internal Verbalization factor will activate phonological representations from images more quickly and/or to a greater extent than people who score lower on the Internal Verbalization factor. Therefore, we predicted a negative correlation between internal verbalization and reaction times (RTs), particularly on trials on which the cue and the target match (we anticipated accuracy would be at ceiling, and so our main outcome variable is RTs). If higher internal verbalization is associated with more automatic/robust phonological activation, then it should also result in greater phonological interference, e.g., slower RTs in rejecting a picture of a foot after a phonologically related cued such as "root". Lastly, we predicted that internal verbalization will be associated with reduced semantic interference: people with higher internal verbalization should show less semantic interference, i.e., be less slowed by more semantically related cue-target pairs.

This last prediction stems directly from the label feedback hypothesis (Lupyan, 2012a, 2012b) according to which perceptual inputs that have been previously associated with labels (true of all the materials we use in the present task) will automatically activate the associated label. The label then feeds back and helps activate category diagnostic features. For example, the word "foot" is selectively associated with visual features that help distinguish feet from semantically related concepts such as shoes and legs. Although these were our main a priori predictions, we also examined the relationship between performance and the other two imagery-related factors: visual and orthographic imagery, and the representational manipulation factor. These analyses should be viewed as exploratory.

Methods

Participants

We recruited undergraduate students from the University of Wisconsin-Madison psychology participant pool who completed the IRQ as part of a larger survey at the start of the semester, targeting participants who passed the attention checks, were fluent English speakers (seven were non-native English speakers) and were below the 25% percentile or above the 75% percentile on the verbal factor of the IRQ. We tested 56 students, and excluded any participants for having an accuracy rate below 90%. Participants were between 18



Fig. 2 Example cue-target match trials (top) and cue-target mismatch trials (bottom). The timing parameters were identical for match and mismatch trials

and 31 years old (21 male, 34 female, M_{age} =18.82 years; SD = 1.83).

Materials

To test the hypothesis that IRQ scores predict sensitivity to different forms of cue-target similarity, we selected words that were related to one another on phonological, orthographic, and semantic dimensions. We recruited an additional 37 participants (UW-Madison undergraduates). Each participant rated 70 word-pairs on phonological, orthographic, and semantic similarity on a 1–7 Likert-style scale ranging from 1: *Completely Different*, to 7: *Identical*. We included disambiguated homonym pairs e.g., bat (the animal); bat (for sport) as attention checks. The homonyms ensured that there were items that should be rated as identical for phonology and orthography, but different for semantic similarity. Five participants were excluded from the rating task for failing the attention checks.

Participants received an example to highlight different types of similarity: "the words cold (for weather) and cold (for illness), mean quite different things, but they look and sound identical. The words lint and tint both look and sound quite similar if they were said aloud. However, words that look similar don't always sound similar e.g., lint and pint." Participants were asked to rate each word pair on three scales: Meaning (what the two words relate to) Sound (how the words sound when said aloud) Look (the visual appearance of the words)

On each trial, participants saw a word pair, e.g., "footroot". Every participant rated each word pair on the three scales above. The results are shown in Table 4 and form the similarity measures used in the main analyses presented below.

Speeded verification task procedure

Each participant completed 288 trials. Participants in the picture-to-text condition had to indicate whether a text word (the target) matched a preceding picture (the cue). The text-topicture condition was analogous with participants indicating whether a picture target matched the preceding text word acting as a cue. The cues and targets were 36 monosyllabic words naming familiar animals and artifacts (see Table 4). We created four exemplars of each word in each modality, i.e., four different pictures of the same object and four text exemplars (lower and upper case, and fonts "Times New Roman", and "Courier New"). The word pairs were equated on concreteness and word frequency (Brysbaert, Warriner, & Kuperman, 2014) sensory experience ratings (Juhasz & Yap, 2013) and imageability based on the norms from the MRC psycholinguistic database (Coltheart, 1981). Of the non-match trials, 50% of the presented cues and targets did not rhyme or share similarities in spelling, e.g., "clock" and "whale". The rest of

Pair	Orthographic similarity	Phonological similarity	Semantic similarity
bear hair	2.66	5.28	1.19
bomb watch	1.06	1.12	1.22
bowl owl	5.34	2.72	1.09
box fox	5.41	5.59	1.16
box tree	1.06	1.03	1.00
cake hat	1.12	1.12	1.03
cake rake	5.69	5.75	1.06
cat hat	5.44	5.78	1.09
chair foot	1.03	1.12	1.25
clock bear	1.09	1.06	1.03
comb bomb	5.62	2.50	1.03
comb train	1.09	1.09	1.00
cow bow	5.62	3.78	1.09
deer beer	5.75	5.84	1.09
deer hair	1.44	1.75	1.06
fox shoe	1.16	1.03	1.00
key tree	2.38	4.91	1.09
owl key	1.12	1.00	1.00
pear plane	2.12	1.91	1.06
plane train	2.50	5.41	2.44
rake root	2.12	2.09	1.25
root foot	5.72	3.88	1.50
rope cow	1.38	1.03	1.06
shell bell	4.88	5.44	1.16
shell bow	1.03	1.06	1.09
shoe toe	4.44	1.81	1.81
snail beer	1.00	1.03	1.00
snail whale	2.34	5.38	1.25
soap match	1.06	1.06	1.00
soap rope	3.00	5.38	1.06
sock bell	1.12	1.09	1.03
sock clock	4.81	5.44	1.03
toe bowl	1.44	2.59	1.00
watch match	5.62	3.50	1.16
whale cat	1.06	1.06	1.28

 Table 4.
 Mean ratings for orthographic, phonological, and semantic similarity for each word pair (1= Completely different; 7 = Identical)

the non-match pairs were randomized to either orthographically rhyme (the rhyme is congruent with the spelling) e.g., "rake" and "cake"; non-orthographically rhyme (the rhyme is in-congruent with the spelling) e.g., "whale" and "snail"; or words that were spelt similarly but did not rhyme e.g., "match" and "watch".

Each trial began with a cue (text or picture, depending on condition) presented for 500 ms. Following a jittered delay (800–1200ms, in 100-ms increments), the target appeared and remained visible until a button response (match, no-match). An incorrect response elicited a buzzing sound and a 1-s timeout. The trials were divided into *match* trials (144

trials; 50%) and *non-match* trials on which the cue-target similarity was varied to investigate the contributions of the variables shown in Table 4.

Results

Data analysis

Mean accuracy was high (98%), with subject-means ranging from 92% to 100%. Faster reaction times were associated with higher accuracy on match trials; there was no relationship

between RTs and accuracy on non-match trials. These findings suggest that there was no speed–accuracy tradeoff. Our analysis therefore focuses on correct RTs (see supplementary materials for full reporting of accuracy). We excluded 5% of all trials for RTs that were too short (< 150 ms) excessively long (>1500 ms). All analyses were conducted in R using mixed effects models with subject and cue (e.g., foot, root, etc.) as random effects using the lme4 package (Bates, Mächler, Bolker, & Walker, 2014).

Effects of cue-target relationships on verification times

Before reporting how the IRQ relates to people's performance, we report here how the relationships between the cue and the target affect verification responses for the sample as a whole by examining (1) the effects of cue-target type: picture cue \rightarrow text target vs. text cue \rightarrow picture target, and (2) the type of similarity between the cue and target, i.e., semantic, phonological, and orthographic. In English, words that are orthographically similar tend to be also phonologically similar, but owing to irregularities of the English writing system, the two factors can be dissociated (e.g., "root" and "foot" share a vowel orthographically, but not phonologically). In our sample, the correlation between orthographic and phonological similarity ratings was.71. To account for independent contributions of phonology and orthography, we residualized phonology on orthography, and orthography on phonology in the model. The full model (using centered and scaled predictors and including mismatching trials only) was:

$$\label{eq:loss_relation} \begin{split} & \texttt{nonmatch_RT}-\texttt{phonological_similarity} \texttt{*cue_type} + \texttt{orthographic_similarity} \texttt{*cue_type} + \texttt{semantic_similarity} \texttt{*cue_type} + (1|\texttt{participant}) + (1|\texttt{cue}) \end{split}$$

The effects of different similarity types on RTs are shown in Fig. 3. When the cue and target did not match, participants were slower to respond if the cue and target were orthographically similar, b = 8.65 (SE = 2.11), t = 4.09. Phonological similarity between the cue and target did not significantly predict RTs b = 0.84 (SE = 2.20), t = 0.38. We also observed semantic interference: participants were slower to respond to confirm that the target did not match the cue when the cue and target were more semantically related, b = 10.64 (SE = 2.20), t = 4.84.

Cue-target type (picture cue followed by text target, vs. text cue followed by picture target) did not significantly predict overall RT b = 17.20 (SE = 16.19), t = 1.06. However, there was a significant interaction between orthographic similarity and cue-target type b = -10.95 (SE = 2.10), t = -5.22, when cued by text, greater orthographic similarity between the text cue and the picture target predicted slower RTs, when cued by a picture, orthographic similarity did not predict RTs. There was no interaction between phonological similarity and cuetarget type b = 1.50 (SE = 2.18), t = 0.69.

Do the IRQ factors predict speed of response for *matching* trials?

We first examined whether the IRQ predicted RTs on matchtrials (e.g., seeing "cat" followed by an image of a cat, or seeing an image of a cat followed by the word "cat"). Results are shown in Fig. 4. The full model (using *z*-scored predictors) was:

matchRT~IRQ_factor*cue_type + (1|participant) + (1|cue)

where cue_type is a centered predictor coding whether the cue was an image (-.5) or text (.5), and cue is cue category (e.g., the word "foot" or a picture of a foot).

Internal verbalization Participants who had higher Verbal IRQ scores responded to match trials more slowly when the cue was a picture: b = 78.80 (SE = 21.86), t = 3.60, but not when the cue was text, b = -40.78 (SE = 33.41), t = -1.22, yielding a significant cue-type by verbal factor interaction, b = -61.94 (SE = 19.73), t = -3.14.

Visual imagery Visual imagery did not predict RTs when the cue was a picture b = -22.57 (SE = 27.72), t = -0.81, but did predict RT's when the cue was text b = 76.72 (SE = 33.17), t = 2.31, yielding a significant interaction with cue-type b = 47.59 (SE = 23.09), t = 2.06.

Orthographic imagery Orthographic imagery did not predict RTs either when the cue was a picture b = -13.28 (SE = 23.25), t = -0.57, or when the cue was text b = 42.21 (SE = 36.73), t = 1.15.

Manipulational representation Greater scores on the manipulational representation factor predicted faster overall RT b = -53.24 (SE = 20.50), t = -2.60 regardless of cuetarget type (there was no interaction between manipulational representation and cue-target type) b = -23.38 (SE = 20.22), t = -1.16.

Do the IRQ factors predict speed of response for *mismatching* trials?

Analyzing the relationship between IRQ scores and performance on the mismatching trials allowed us to determine whether people's IRQ profiles were related to different types of interference between the cue and the target. Results are shown in Fig. 5. The full model syntax (using centered and scaled predictors) was:

 $orthographic_similarity + IRQ_factor*semantic_similarity + (1|participant) + (1|cue)$

 $non-match_RT \sim IRQ_factor*phonological_similarity + IRQ_factor*$



Fig. 3 a The effect of cue-target similarities on mismatch RTs (the time to reject the target as mismatching the cue). Error bands signify + 1/-1 SEs of model-predicted means. RTs are most affected by semantic similarity

when a picture target is matched to a text cue (a) and by orthographic similarity when a text target is matched to a picture cue (b)



Fig. 4 Regression coefficients from mixed effects models predicting cue and target match RT showing the main effects of visual imagery internal verbalization, orthographic imagery, and representational manipulation. *Error bars* show 95% CI of regression coefficients



Fig. 5 Regression coefficients from mixed effects models showing the main effects of visual imagery, internal verbalization, orthographic imagery and representational manipulation as well as their interactions with the three types of cue-target similarity (semantic, orthographic,

phonological). For example, B. shows that greater internal verbalization was associated with a significantly smaller effect of semantic interference, but a larger influence of phonological similarity. *Error bars* show 95% CI of regression coefficients

Internal verbalization Participants with higher internal verbalization scores responded more slowly overall b = 36.08 (SE = 15.38), t = 2.35. This effect was driven by the picture-cue condition: b = 67.67 (SE = 21.16), t = 3.20. When the cue was text, RTs were not predicted by people's propensity to internally verbalize b = 2.20 (SE = 21.67), t = 0.10. The cue-type by internal verbalization interaction was reliable, b = -35.76 (SE = 15.50), t = -2.31. This interaction is visualized in Fig. 6. As the figure shows, while participants with lower internal verbalization scores were faster on picture-to-text trials (dashed lined in Fig. 6), participants higher on the internal verbalization factor were not. Indeed, for participants with the highest propensity to internally verbalize, it was text-to-picture trials that were faster.

There was no *overall* effect of phonological similarity on RTs, greater internal verbalization predicted slower RTs when the cue and target were phonologically similar (i.e., greater phonological interference), b = 4.69 (SE = 2.21), t = 2.13. The relationship between phonological similarity and internal verbalization was similar for text-to-picture and picture-to-text trials, b = 3.21 (SE = 2.22), t = 1.44.

The orthographic similarity of the cue and the target did not significantly interact with internal verbalization b = 2.27 (SE = 2.12), t = 1.07.

Although there was an overall effect of semantic interference, those with a higher propensity for internal verbalization showed reduced semantic interference on textto-picture trials, b = -8.50 (SE = 3.07), t = -2.77. On picture-to-text trials, the effect of semantic similarity on RTs was not related to differences in internal verbalization, b = 1.24 (SE = 2.99), t = 0.41, leading to a significant interaction between cue-target type and internal verbalization, b = -5.80 (SE = 2.15), t = -2.70. **Visual imagery** Visual imagery did not predict overall RTs, but participants with higher visual imagery scores responded more slowly if the cue and the target were orthographically similar b = 5.40 (SE = 2.19), t = 2.47. The effect did not interact with cue-target type, the relationship between orthographic similarity and visual imagery was consistent whether the cue was a text and the target a picture, or when the cue was a picture and the target was text b = -2.32 (SE = 2.22), t = -1.04. Visual imagery did not interact with either phonological or semantic similarity (p > .05).

Orthographic imagery Orthographic imagery did not predict overall RTs, but participants who showed a greater propensity for orthographic imagery responded more slowly if the cue and the target shared orthographic similarity b = 6.68 (SE = 2.10), t = 3.18. There was no interaction with cue-target type; the effect of orthographic similarity and orthographic imagery was consistent whether the cue was text and the target a picture, or when the cue was a picture and the target was text b = -2.58 (SE = 2.11), t = -1.22. Orthographic imagery did not significantly interact with phonological or semantic similarity (p > .05).

Representational manipulation did not predict RTs for mismatching trials b = -14.02 (SE = 16.39), t = -0.86. There was no interaction with cue-target type, b = 7.49 (SE = 16.46), t = 0.46. Representational Manipulation did not interact with phonological or semantic similarity (p > .05).

General discussion

Many people feel that their thinking takes the form of language, often in the form of a conversation with oneself. Others rarely experience this, or deny having this experience altogether. For example, 19% of respondents disagreed with the statement "I hear words in my 'mind's ear' when I think" and 16% disagreed with the statement "I think about problems in my mind in the form of a conversation with myself".

We designed the Internal Representation Questionnaire (IRQ) to measure differences in people's subjective experience of their thoughts. The internal verbalization factor of the IRQ shows substantial correlations with several previously developed inner-speech measures, particularly the Dialogic component of the varieties of Inner Speech Questionnaire (VISQ; McCarthy-Jones & Fernyhough, 2011; Alderson-Day et al., 2018) and the Management component of the Self Talk Scale (STS; Brinthaupt et al., 2009), see Table 3 (cf. Uttl et al., 2011). One difference between how we operationalize internal verbalization and the way it has tended to be operationalized in work on inner speech is that we view language as a control system, augmenting "nonverbal" computations rather than as a separate representational medium. For example, an endorsement of an item such as "If I am walking somewhere by myself, I often have a silent conversation with myself' may connote a person's greater habitual engagement of language as input to such nonlinguistic processes as visual imagery.

One advantage of the IRQ over other questionnaires assessing internal verbalization is that its 36 questions also assess vividness of visual imagery (drawing on questions from Kirby et al., 1988: Blajenkova et al., 2006: Marks, 1973), and what appears to be a more general representational manipulation factor. This factor arose from our inclusion of questions probing static visual imagery (e.g., focusing on objects and faces), and dynamic (spatial) visual imagery (Blajenkova et al., 2006). Questions probing dynamic visual imagery indeed clustered separately from questions probing static imagery, but clustered together with questions probing imagery in other modalities, e.g., dynamic auditory imagery: "I can easily choose to imagine this sentence in my mind pronounced unnaturally slowly" and tactile imagery: "It is easy for me to imagine the sensation of licking a brick". The positive relationship of this factor to Need For Cognition is intriguing and in need of further investigation.

Finally, our initial inclusion of a variety of imagery-focused questions reveals what appears to be a previously undescribed orthographic imagery factor. For example, 20% of participants agreed with the statement "When I hear someone talking, I see words written down in my mind" and 36% agreed with the statement "I see words in my "mind's eye" when I think".

As an initial test of the IRQ's predictive validity, we used it to predict people's performance on a speeded cue-target verification task. On each trial participants either saw a written text cue followed by a picture target, or a picture cue followed by a written text target. In both cases, they had to indicate, as quickly as possible, whether the target and cue matched. On mismatching trials, we systematically manipulated the relationship between the cue and the target on phonological



Mismatching cue and target



Fig. 6 a The effect of internal verbalization and cue-target type on match and mismatch RTs. Error bands signify + 1/-1 SEs of model-predicted means. RTs are affected by internal verbalization when the cue is a picture

and the target is a text word for both matching and mismatching cuetarget types, but not for text cues-picture targets

similarity, orthographic similarity, and semantically similarity (e.g., shoe and toe).

If people with a greater internal verbalization propensity as revealed by higher Internal Verbalization scores on the IRQ are more likely to name images and/or activate phonological representations more quickly/robustly from written text, we expected to find greater phonological interference (i.e., a steeper slope of the green line in Fig. 3a) when the cue and target were phonologically related (e.g., the word "soap" followed by a picture of a rope). This prediction was confirmed: people with greater Internal Verbalization scores on the IRQ showed greater phonological interference, but limited to trials in which a text cue is compared to a pictorial target (Fig. 5b, f). This finding extends earlier work showing that people who score higher on the verbal dimension of the Verbal-Visual Questionnaire (Kirby et al., 1988), show greater activation in the left supramarginal gyrus (Brodmann's area 40) (Kraemer et al., 2009), thought to be linked to phonological processing (e.g., Smith & Jonides, 1998).

To the extent that language emphasizes categorical distinctions (Forder & Lupyan, 2019; Lupyan, 2012b), with labels selectively activating category-diagnostic features (Lupyan & Thompson-Schill, 2012), we expected that a greater reliance on language in the cue-target verification task would lead to less semantic interference. Although the cues and targets had relatively little semantic overlap (see Table 4), we observed a robust semantic interference effect (the blue line in Fig. 3a). For example, participants were slower to respond "mismatch" when a text cue (e.g., "shoe") was semantically related to a picture target (e.g., toe). Consistent with our prediction, participants with higher internal verbalization were less affected (less slowed) by trials where a text cue was semantically related to a picture target.

To our surprise, higher internal verbalization was associated with overall slower RTs (with no evidence of a speedaccuracy tradeoff), specifically when participants matched a picture cue to a text target. Suppose that participants perform this task by naming the picture and matching the name to the written word. If the propensity to internally verbalize measure is associated with greater phonological activation (as suggested by the phonological interference result summarized above), it is puzzling why it is associated with slower performance. A post hoc account of what may be happening is to consider the cue-target task from a perspective in which the cue sets up an expectation (a prior) within which the target word/image is then processed (Boutonnet & Lupyan, 2015). Shorter RTs indicate that the prior facilitated the processing of the target word or image. Viewed in this way, the slower responding of people with greater Internal Verbalization scores-specifically when pictures are used as cuessuggests that these participants are less efficient in using pictorial cues. Notably, the text cue provides a concrete prior on which to activate phonology, in a way that a picture does not. Our design does not allow us to determine whether the RT cost of pictorial cues extends to processing different kinds of targets, e.g., matching a picture to a picture from the same or different category. Turning to the more exploratory findings: when the cue was a picture and the target a written word, we observed an orthographic interference effect-a slowing of responses when the name of the picture and the target word had similar orthography (adjusting for phonology) (cf. Walenchok, Hout, & Goldinger, 2016; Barca, Benedetti, & Pezzulo, 2016; Zelinsky & Murphy, 2000). This orthographic interference effect is shown in red in Fig. 3b. The strength of this orthographic interference (steepness of the red line) was correlated with visual and orthographic IRQ factors (see Fig. 5e, g). This finding is consistent with the possibility that people scoring higher on these factors are activating orthographic representations from pictures to a greater extent.

Although the overall pattern of findings from the cue-target verification study, as it relates to people's IRQ profiles is rather complex, the effects suggest that differences in the IRQ do predict certain behavioral differences. We view the current cue-target paradigm as just the first step in linking differences in IRQ profiles to differences in behavior.

Relationship between mode of thinking and language-augmented thought

On one still widespread view of the relationship between language and cognition is that language is largely a vehicle for communicating our thoughts. With this view, language is thought to play a minor role in the construction of thoughts (if any role at all) (Devitt & Sterelny, 1987; Li & Gleitman, 2002; McWhorter, 2014). On another view, language is viewed as a separate representational medium. From this view, one can "think in images" or "think in words" with the two modes of thought occurring in distinct modalities (Carruthers, 2002). We endorse another position, on which language has the power to augment non-linguistic cognitive and perceptual processes (Clark, 1998; Lupyan, 2012a, 2012b for discussion, 2016). From this view, language is not a distinct representational medium, but rather a mechanism to augment mental representations into a more categorical form, which promotes reuse and compositionality. In support of this perspective, mental representations elicited by language are more categorical than those elicited by informationally equivalent nonverbal cues (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012), and these differences can be seen even in lowerlevel perceptual tasks. For example, hearing a color word temporarily causes people to perceive colors in a more categorical way, changing patterns of discrimination accuracy (Forder & Lupyan, 2019). This prior work shows that up-regulating language, e.g., through overt presentation of verbal labels, facilitates categorization, while interfering with language disrupts it. For example, verbal interference affected people's ability to categorize pictures according to common perceptual attributes, e.g., grouping a snowman and a swan together on a basis of a shared color (Lupyan, 2009).

A key prediction of this view as it relates to individual differences in internal verbalization is that people whose thinking feels to them as more language-like may show more categorical processing on a variety of domains such as reasoning, mental imagery, and patterns of errors in recall.

Studying the relationships between IRQ profiles and performance on other tasks can help us understand how differences in phenomenology relate to differences in behavior. For example, to the extent that it is easier to align more categorical mental representations, do people with greater internal verbalization show greater alignment in their mental representations? In some past work, we have shown that verbal interference disrupts certain types of categorization (suggesting that language is ordinarily involved in such categorization) (Lupyan, 2009). Do people who report higher levels of inner verbalization show greater disruption-suggesting that they are more reliant on language), less disruption-suggesting that their language-augmented processing is more robust to interference, or is it unrelated-suggesting that even people who do not experience inner verbalization rely on language to the same extent, but this reliance is not accompanied by conscious experience.

Limitations

Any instrument attempting to quantify subjective experience is limited by people's self-report. Most of the statements included in the IRQ probe trait-like qualities, e.g., "I hear words in my "mind's ear" when I think"). Some items require participants to retrospect about their experience of specific situations, e.g., "If I talk to myself in my head, it is usually accompanied by visual imagery"). Although people's answers to these questions appears to be quite stable as judged by high test-retest reliability, we do not know how well their responses to these questions track in-the-moment subjective experience. One way to find out is to correlate IRQ scores with results obtained from experience sampling methods. One such method is Descriptive Experience Sampling (DES; Hurlburt and Akhter (2006)): participants wear a beeper and when it beeps (typically six times per day) are asked to attend to "whatever was directly present, ongoing in their inner experience the microsecond before the beep began, and to jot down notes about that experience." The objective is to probe the "last undisturbed moment of pristine inner experience before the beep" (Hurlburt et al., 2013, p. 1479). Some proponents of such sampling methods take a somewhat combative stance regarding the use of questionnaires (Hurlburt et al., 2013). While there is no doubt that experience sampling is capable of producing a much richer report of the quality of an individual's experience as compared to a questionnaire, the

interesting question is the extent to which the phenomenology revealed by sampling and questionnaire methods aligns with people's questionnaire responses (Alderson-Day & Fernyhough, 2015b). When it does not (Alderson-Day & Fernyhough, 2015b), we can try to understand why. An interesting possibility is that sampling methods are more sensitive about habitual in-the-moment experiences while questionnaire responses are more sensitive to people's control over their experiences. For example, someone who can more easily induce the experience of inner speech may, when responding to a questionnaire, over-estimate the extent to which they actually experience it moment-to-moment. Which of these is more important in understanding the relationship between differences in phenomenology and differences in objective behavior, remains an open question.

Conclusions

We introduce a new instrument, the Internal Representation Questionnaire (IRQ). The primary motivation for the IRQ is to measure the extent to which people experience their thoughts in the form of language and use language to guide their thinking (glossed here as internal verbalization). To increase its usefulness, we included in the IRQ items measuring vividness of visual imagery, items pertaining to specifically orthographic imagery-a construct that, to our knowledge, has not been previously described-and items that measure people's subjective ease of manipulating mental representations across different modalities. The IRQ has high internal validity and good test-retest reliability. We presented one test of its predictive validity by using people's IRQ profiles to predict performance on a speeded cue-target verification experiment (Fig. 2). This validation confirmed some of our predictions: people who internally verbalize will show less semantic interference and greater phonological interference. Counter to our prediction those internally verbalizing to a greater extent responded more slowly when matching a text target to a picture cue suggesting that although people with higher internal verbalization scores also report greater use of visual imagery, they may be less efficient in using pictorial cues. We also observed that people with greater visual and orthographic imagery were more sensitive to orthographic similarity between the cue and the target (i.e., they showed more interference when the target word "root" was preceded by a picture of a foot). This validation of the IRQ is just one step toward understanding the relationship between the phenomenology of thought and its relationship with objective performance. People seem to show large differences in how the experience their thoughts. We are optimistic about the usefulness of the IRQ in uncovering the objective consequences of these differences.

Open Practices Statement The data and materials are available at https://osf.io/8rdzh/. This experiment was not preregistered.

References

- Alderson-Day, B., & Fernyhough, C. (2015a). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, 141 (5), 931–965. doi:https://doi.org/ 10.1037/bul0000021
- Alderson-Day, B., & Fernyhough, C. (2015b). Relations among questionnaire and experience sampling measures of inner speech: A smartphone app study. *Frontiers in Psychology*, 6. doi:https://doi. org/10.3389/fpsyg.2015.00517
- Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The varieties of inner speech questionnaire (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition*, 65, 48–58. doi: https://doi.org/10.1016/j.concog.2018.07.001
- Amedi, A., Malach, R., & Pascual-Leone, A. (2005). Negative BOLD Differentiates Visual Imagery and Perception. *Neuron*, 48 (5), 859– 872. doi:https://doi.org/10.1016/j.neuron.2005.10.032
- Baldo, J. V., Paulraj, S. R., Curran, B. C., & Dronkers, N. F. (2015). Impaired reasoning and problem-solving in individuals with language impairment due to aphasia or language delay. *Frontiers in Psychology*, 6. doi:https://doi.org/10.3389/fpsyg. 2015.01523
- Barca, L., Benedetti, F., & Pezzulo, G. (2016). The effects of phonological similarity on the semantic categorisation of pictorial and lexical stimuli: Evidence from continuous behavioural measures. *Journal of Cognitive Psychology*, 28 (2), 159–170. doi:https://doi.org/10.1080/ 20445911.2015.1101117
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. Retrieved from http://arxiv.org/ abs/1406.5823
- Beyler, J., & Schmeck, R. R. (1992). Assessment of Individual Differences in Preferences for Holistic-Analytic Strategies: Evaluation of Some Commonly Available Instruments. *Educational and Psychological Measurement*, 52 (3), 709–719. doi:https://doi.org/10.1177/0013164492052003021
- Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology*, 20 (2), 239–263. doi:https://doi.org/10. 1002/acp.1182
- Boutonnet, B., & Lupyan, G. (2015). Words Jump-Start Vision: A Label Advantage in Object Recognition. *Journal of Neuroscience*, 35 (25), 9329–9335. doi:https://doi.org/10.1523/JNEUROSCI.5111-14. 2015
- Bowerman, M., & Levinson, S. C. (2001). Language acquisition and conceptual development. Cambridge University Press.
- Brewer, W. F., & Schommer-Aikins, M. (2006). Scientists Are Not Deficient in Mental Imagery: Galton Revised. *Review of General Psychology*, 10 (2), 130–146. doi:https://doi.org/10.1037/1089-2680.10.2.130
- Brinthaupt, T. M., Hein, M. B., & Kramer, T. E. (2009). The Self-Talk Scale: Development, Factor Analysis, and Validation. *Journal of Personality Assessment*, 91 (1), 82–92. doi:https://doi.org/10.1080/ 00223890802484498
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46 (3), 904–911. doi:https://doi.org/ 10.3758/s13428-013-0403-5
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. Journal of Personality and Social Psychology, 42 (1), 116–131. doi:https://doi. org/10.1037/0022-3514.42.1.116
- Calvete, E., Estévez, A., Landín, C., Martínez, Y., Cardeñoso, O., Villardón, L., & Villa, A. (2005). Self-Talk and Affective Problems in College Students: Valence of Thinking and Cognitive

🙆 Springer

Content Specificity. *The Spanish Journal of Psychology*, 8 (01), 56–67. doi:https://doi.org/10.1017/S1138741600004960

- Carruthers, P. (2002). The cognitive functions of language. *Behavioral* and Brain Sciences, 25 (6), 657–674. doi:https://doi.org/10.1017/ S0140525X02000122
- Clark, A. (1997). Being there: Putting brain, body, and world together again. Cambridge, Mass: MIT Press.
- Clark, A. (1998). Magic Words: How language augments human computation. In Language and thought: Interdisciplinary themes (pp. 162– 183). New York, NY: Cambridge University Press.
- Clark, L. A., & Watson, D. (1995). Constructing Validity: Basic Issues in Objective Scale Development. *Psychological Assessment*, 7 (3), 309–19.
- Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology Section A*, *33* (4), 497–505. doi:https://doi.org/10.1080/14640748108400805
- Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research*, 47 (4), 474–478. doi:https:// doi.org/10.1016/j.visres.2006.11.013
- Devitt, M., & Sterelny, K. (1987). Language and reality: An introduction to the philosophy of language. MIT Press.
- Duncan, R. M., & Cheyne, J. A. (1999). Incidence and functions of selfreported private speech in young adults: A self-verbalization questionnaire. Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement, 31 (2),133–136. doi: https://doi.org/10.1037/h0087081
- Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, 143, 93–100. doi:https://doi.org/10. 1016/j.cognition.2015.06.008
- Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, 148 (7), 1105–1123. doi:https://doi.org/10.1037/xge0000560
- Galton, F. (1880). Statistics of Mental Imagery. Mind, 5 (19), 301-318.
- Hatakeyama, T. (1997). Adults and Children with High Imagery Show More Pronounced Perceptual Priming Effect. *Perceptual and Motor Skills*, 84 (3_suppl), 1315–1329.doi:https://doi.org/10.2466/pms. 1997.84.3c.1315
- Heavey, C. L., & Hurlburt, R. T. (2008). The phenomena of inner experience. *Consciousness and Cognition*, 17 (3), 798–810. doi:https:// doi.org/10.1016/j.concog.2007.12.006
- Heavey, C. L., Moynihan, S. A., Brouwers, V. P., Lapping-Carr, L., Krumm, A. E., Kelsey, J. M., Hurlburt, R. T. (2018). Measuring the Frequency of Inner-Experience Characteristics by Self-Report: The Nevada Inner Experience Questionnaire. *Frontiers in Psychology*, 9, 2615. doi:https://doi.org/10.3389/ fpsyg.2018.02615
- Hill, B. D., Foster, J. D., Elliott, E. M., Shelton, J. T., McCain, J., & Gouvier, W. D. (2013). Need for cognition is related to higher general intelligence, fluid intelligence, and crystallized intelligence, but not working memory. *Journal of Research in Personality*, 47 (1), 22–25. doi:https://doi.org/10.1016/j.jrp. 2012.11.001
- Holmes, K. J., & Wolff, P. (2013). Spatial language and the psychological reality of schematization. *Cognitive Processing*, 14 (2), 205–208. doi:https://doi.org/10.1007/s10339-013-0545-5
- Hurlburt, R. T., & Akhter, S. A. (2006). The Descriptive Experience Sampling method. *Phenomenology and the Cognitive Sciences*, 5 (3-4), 271–301. doi:https://doi.org/10.1007/s11097-006-9024-0
- Hurlburt, R. T., Heavey, C. L., & Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Consciousness and Cognition*, 22 (4), 1477–1494. doi:https://doi.org/10.1016/j. concog.2013.10.003

- Jonassen, D. H., & Grabowski, B. (1993). Individual differences and instruction. *New York: Allen & Bacon.*
- Juhasz, B. J., & Yap, M. J. (2013). Sensory experience ratings for over 5, 000 mono- and disyllabic words. *Behavior Research Methods*, 45 (1), 160–168. doi:https://doi.org/10.3758/s13428-012-0242-9
- kafoBoto. (2016). *Do you hear a voice while you read*? Reddit. Retrieved from https://www.reddit.com/r/books/comments/4wttu4/do_you_hear a voice while you read
- Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 105, 53–60. doi:https://doi.org/ 10.1016/j.cortex.2017.10.012
- Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology*, 13 (2), 169–184. doi:https://doi.org/10.1016/0361-476X(88) 90017-3
- Klinger, E., & Cox, W. M. (1987). Dimensions of Thought Flow in Everyday Life. *Imagination, Cognition and Personality*, 7 (2), 105–128. doi:https://doi.org/10.2190/7K24-G343-MTQW-115V
- Kosslyn, S. M., Brunn, J., Cave, K. R., & Wallach, R. W. (1984). Individual differences in mental imagery ability: A computational analysis. *Cognition*, 18 (1-3), 195–243. doi:https://doi.org/10.1016/ 0010-0277(84)90025-8
- Kraemer, D. J. M., Rosenberg, L. M., & Thompson-Schill, S. L. (2009). The neural 781correlates of visual and verbal cognitive styles. *The Journal of Neuroscience*, 78229 (12), 3792–3798. doi:https://doi.org/10.1523/JNEUROSCI.4635-08.2009
- Kunkletown. (2012). What does your inner voice sound like? Reddit. Retrieved from https://www.reddit.com/r/AskReddit/ comments/p7osx/mute_ people_of_reddit_what_does_your_ inner voice/
- Langland-Hassan, P., Faries, F. R., Richardson, M. J., & Dietz, A. (2015). Inner speech deficits in people with aphasia. *Frontiers in Psychology*, 6. doi:https://doi.org/10.3389/fpsyg.2015.00528
- Leutner, D., & Plass, J. L. (1998). Measuring learning styles with questionnaires versus direct observation of preferential choice behavior in authentic learning situations: The visualizer/ verbalizer behavior observation scale (VV-BOS). *Computers in Human Behavior*, 14 (4), 543–557. doi:https://doi.org/10. 1016/S0747-5632(98)00023-5
- Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, 83 (3), 265–294. doi:https://doi.org/10.1016/ S0010-0277(02)00009-4
- Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review*, 16 (4), 711–718. doi:https://doi. org/10.3758/PBR.16.4.711
- Lupyan, G. (2012a). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology*, 3. doi: https://doi.org/10.3389/fpsyg.2012.00054
- Lupyan, G. (2012b). What Do Words Do? Toward a Theory of Language-Augmented Thought. In *Psychology of Learning and Motivation* (Vol. 57, pp. 255–297). Elsevier. doi:https://doi.org/10.1016/B978-0-12-394293-7.00007-8
- Lupyan, G. (2015). The paradox of the universal triangle: Concepts, language, and prototypes. *Quarterly Journal of Experimental Psychology*. doi:https://doi.org/10.1080/17470218.2015.1130730
- Lupyan, G. (2016). The Centrality of Language in Human Cognition: Centrality of Language. *Language Learning*, 66 (3), 516–553. doi: https://doi.org/10.1111/lang.12155
- Lupyan, G., & Bergen, B. (2016). How Language Programs the Mind. *Topics in Cognitive Science*, 8 (2), 408–424. doi:https://doi.org/10. 1111/tops.12155
- Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means.

Journal of Experimental Psychology: General, 141 (1), 170–186. doi:https://doi.org/10.1037/a0024904

- Macleod, C. M., Hunt, E. B., & Mathews, N. N. (1978). Individual differences in the verification of sentence—picture relationships. *Journal of Verbal Learning and 816 Verbal Behavior*, *17* (5), 493–507. doi:https://doi.org/10.1016/S0022-5371(78) 90293-1
- Marks, D. F. (1973). Visual Imagery Differences in the Recall of Pictures. British Journal of Psychology, 64 (1), 17–24. doi:https://doi.org/10. 1111/j.2044-8295.1973.tb01322.x
- Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preferences. *Journal of Educational Psychology*, 833–846.
- McCarthy-Jones, S., & Fernyhough, C. (2011). The varieties of inner speech: Links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition*, 20 (4), 1586–1593. doi:https://doi.org/10.1016/j. concog.2011.08.005
- McKelvie, S. J. (1994). The Vividness of Visual Imagery Questionnaire as a predictor of facial recognition memory performance. *British Journal of Psychology*, 85 (1), 93–104. doi:https://doi.org/10.1111/ j.2044-8295.1994.tb02510.x
- McKelvie, S. J., & Demers, E. G. (1979). Individual differences in reported visual imagery and memory performance*. *British Journal of Psychology*, 70 (1), 51–57. doi:https://doi.org/10.1111/j.2044-8295. 1979.tb02142.x
- McWhorter, J. H. (2014). *The language hoax: Why the world looks the same in any language*. Oxford University Press.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning Styles: Concepts and Evidence. *Psychological Science in the Public Interest*, 9 (3), 105–119. doi:https://doi.org/10.1111/j.1539-6053.2009.01038.x
- Riding, R., & Cheema, I. (1991). Cognitive Styles—an overview and integration. *Educational Psychology*, 11 (3-4), 193–215. doi: https://doi.org/10.1080/0144341910110301
- Rosseel, Y. (2012). An R package for structural equation modeling. *Journal of Statistical Software*, 48 (2), 1–36.
- Schutte, N. S., & Malouff, J. M. (2007). Dimensions of Reading Motivation: Development of an Adult Reading Motivation Scale. *Reading Psychology*, 28 (5), 469–489. doi:https://doi.org/10.1080/ 02702710701568991
- Siegrist, M. (1995). Inner Speech as a Cognitive Process Mediating Self-Consciousness and Inhibiting Self-Deception. *Psychological Reports*, 76 (1), 259–265. doi:https://doi.org/10.2466/pr0.1995.76. 1.259
- Simms, L. J. (2008). Classical and Modern Methods of Psychological Scale Construction. *Social and Personality Psychology Compass*, 2 (1), 414–433. doi:https://doi.org/10.1111/j.1751-9004.2007. 00044.x
- Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences*, 95 (20), 12061–12068. doi:https://doi.org/10.1073/pnas. 95.20.12061
- Sternberg, R. J., Grigorenko, E. L., & Zhang, L.-F. (2008). Styles of Learning and Thinking Matter in Instruction and Assessment. *Perspectives on Psychological Science*, 3 (6), 486–506. doi:https:// doi.org/10.1111/j.1745-6924.2008.00095.x
- Uttl, B., Morin, A., & Hamper, B. (2011). Are Inner Speech Self-Report Questionnaires Reliable and Valid? *Procedia - Social and Behavioral Sciences*, 30, 1719–1723. doi:https://doi.org/10.1016/j. sbspro.2011.10.332
- Walenchok, S. C., Hout, M. C., & Goldinger, S. D. (2016). Implicit object naming in visual search: Evidence from phonological competition. *Attention, Perception, & Psychophysics*, 78 (8), 2633–2654. doi: https://doi.org/10.3758/s13414-016-1184-6

- Wasserman, E. A., & Castro, L. (2012). Categorical Discrimination in Humans and Animals. In *Psychology of Learning and Motivation* (Vol. 56, pp. 145–184). Elsevier. Retrieved from http://linkinghub. elsevier.com/retrieve/pii/B9780123943934000054
- Zelinsky, G., & Murphy, G. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science*, *11* (2), 125–131.
- Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery Congenital aphantasia. *Cortex*, 73, 378–380. doi:https://doi.org/10. 1016/j.cortex.2015.05.019

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.