Factors Underlying Conceptual Change in the Sciences and Social Sciences

Angele Yazbec (angeleyazbec@gmail.com)
Department of Psychology, 1107 W. Call Street
Tallahassee, FL 32306

Arielle Borovsky (aborovsky@purdue.edu)
Department of Speech, Language, and Hearing Sciences
715 Clinic Drive, West Lafayette, IN 47906

Michael Kaschak (kaschak@psy.fsu.edu)
Department of Psychology, 1107 W. Call Street
Tallahassee, FL 32306

Abstract

Learning in the sciences is difficult for students from elementary school to university due to misconceptions, or incorrect prior knowledge, interfering with the acquisition of new knowledge. The process of replacing previously incorrect ideas with new and accurate ones is referred to as conceptual change. Which factors and to what extent they facilitate the conceptual change is debated. This study primarily investigates two key components to conceptual change in scientific knowledge: text style and epistemic beliefs. We also explored additional contributions of individual differences in prior knowledge, reading ability, and working memory. 157 college students completed a two-part, within subjects design study in which they completed pretests, read passages addressing a misconception, completed posttests, and were assessed on a battery of the individual difference measures. We noted conceptual change on the posttest, but individual readers appeared to respond to the text differently.

Keywords: conceptual change; epistemic beliefs; discourse processing

Misconceptions and Conceptual Change

Students may struggle to learn scientific concepts due to pre-existing misconceptions that contradict scientific evidence on how the world operates. Such ideas can be resistant to change because they are strongly held beliefs about the world (e.g. Hewson & Hewson, 1984). Whereas prior knowledge is often useful in comprehending text (e.g. Kintsch, 1988), misconceptions contain inaccurate information, thus creating conflict with educational text. The process of updating these misconceptions with new, more accurate information is referred to as conceptual change (e.g. van den Broek & Kendeou, 2008). According to the Conceptual Change Model, a reader must recognize that (1) the misconception is inadequate for explaining scientific phenomena, (2) the new ideas are intelligible, (3) plausible, and (4) useful for explaining phenomena (Posner, Strike, Hewson, & Gertzog, 1982). Many studies explore the factors that influence conceptual change, but the mechanisms of this process are not entirely clear.

According to the Knowledge Revision Components Framework, a key step to encouraging conceptual change (also referred to as knowledge revision) is making the reader explicitly aware there is a conflict between previously acquired knowledge and new information (Kendeou & O'Brien, 2014). Successful conceptual change occurs when readers integrate new information into their knowledge structure, thus decreasing activation of the debunked misconception.

One way to encourage conceptual change is to read text. Researchers have explored whether refutation texts are better at inducing conceptual change than expository texts. Consider an example targeting the misconception that meteors that land on Earth are hot. Refutation texts start with a preface explicitly stating a misconception such as, "*Kate warned everyone not to touch the meteor because it would be hot and they could get burned. However, Jerry said that they should not worry because it actually should not be hot*." (Kendeou, Walsh, Smith, & O'Brien, 2014, p. 396). Expository text, on the other hand, does not explicitly state the misconception (e.g. "*Kate was excited and curious because she had never seen a meteor on the ground before. Jerry said he could look up more about meteors in the astrophysics book that he had.*") (Kendeou et al., 2014, p. 396). Many studies suggest refutation texts are better at inducing conceptual change than expository texts among school-aged and college students (Tippett, 2010 for review); however, there is also evidence suggesting the two text Styles do not differ in inducing conceptual change (e.g. Hynd & Guzzetti, 1998).

Epistemic beliefs, a person's beliefs about the nature of knowing and the process of knowing, may also affect conceptual change (Hofer & Pintrich, 1997). For example, some people have less advanced epistemic beliefs (e.g. they believe knowledge is rigid and static), thus they are more resistant to changing their knowledge base. Others have more advanced epistemic beliefs (e.g. they believe knowledge is flexible and dynamic), thus they are more willing to change their knowledge base. Whereas more experienced students may readily revise their knowledge in light of reputable evidence, young students may not recognize that their current misconceptions are inadequate for understanding scientific phenomena in the formal education setting (e.g. Posner et al., 1982).

Present Experiment

Ample evidence suggests misconceptions present challenges to learning in the sciences. However, there is conflicting evidence as to whether refutation text is better than expository text in inducing conceptual change. Furthermore, much of this evidence has examined the impact of text style on conceptual change using a between-subjects design (e.g. Lassonde, Kendeou, & O'Brien, 2016), which limits knowledge of individual differences in how much conceptual change readers undergo after reading each text style. There is also some limited evidence suggesting that epistemic beliefs may also play a role in conceptual change. Therefore, we measure the impact of text style and epistemic beliefs on conceptual change using a within subjects study manipulating the style of text participants read (refutation versus expository), which will inform how individual readers differentially respond to different text Styles. If there is an effect of text style and epistemic beliefs on conceptual change, then conceptual change will be greater after participants read refutation texts compared to after they read expository texts. Additionally, we expect that participants with more advanced epistemic beliefs will show greater amounts of conceptual change than students with less advanced epistemic beliefs.

A secondary purpose of this study is to explore how individual differences (e.g. working memory, reading ability, prior knowledge) play a role in conceptual change. Previous findings have suggested that these factors matter for conceptual change (e.g. van den Broek & Kendeou, 2008), while other studies suggest cognitive abilities do not matter (e.g. Toplak, West, & Stanovich, 2014). Thus, the extent to which these factors matter in relation to text style and epistemic beliefs is still not clear but will nevertheless be explored in the present study.

Method

Participants

One hundred and seventy two college-aged participants were recruited from Florida State University's Department of Psychology (*M*= 19.28 years, male= 37). 157 participants completed both sessions and were included in the final sample. The target initial sample size was 100 college-aged adults, but since some of the hypothesized effects were not clearly absent or present according to the prescribed guidelines (see Frick, 1998), we continued data collection until we reached approximately the final cutoff sample size.

Materials

Text Design

Following a similar procedure outlined by Lassonde and colleagues (2016), refutation and expository texts were written for each topic. Five disciplines were tested (Physics, Astronomy, Genetics, Economics, and Geography) and each discipline had two topics that were tested (see Table 1). All text contained the following sections: introduction, premise, explanation, and conclusion. The key difference between the refutation and expository versions of text for each topic was the manipulation of the premise section. The premise section for the refutation text explicitly states the misconception and notes the misconception is incorrect. The premise section for the expository section presents further information about the topic, but does not explicitly state the misconception (see Table 2 for an example). We carefully controlled the length of each text and its subsections. Across all texts, there was a range of 346-349 words. The premise section for each version of text was within one word of each other. The introduction, explanation, and conclusion sections were exactly the same for each text version.

Table 1: List of disciplines and topics assessed in the study.

Discipline	Topic
Physics	Newton's 1st law Newton's 3rd law
Astronomy	Star formation Star colors
Genetics	Alleles Chromosome
Economics	Inflation Gambler's fallacy
Geography	World geography Floridian geography

Table 2: Example premise sections of experimental text.

Expository	Refutation
Whether or not blue stars or red stars are the hottest is not that clear to the casual stargazer. Close investigation of the visible light spectrum of stars, however, does reveal vast temperature differences. Astronomers have devised measures for quantifying the colors of light the stars give off and then using those colors to determine stellar temperatures.	Many people think red stars are hotter than blue stars because red is often associated with fire and other notably hot surfaces or objects. Astronomers, however, have found this idea to be inaccurate. Quantitative measures of starlight color and stellar temperature reveal that blue stars are actually the hottest stars while red stars are the coolest stars.

Pretests and Posttests

Pretests and posttests were identical for each topic. Test questions for the physics topics were pulled from the Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992) and test questions for the astronomy topics were pulled from the Star Properties Concept Inventory (Bailey, Johnson, Prather, & Slater, 2011). Items for the genetics questions were taken from the Genetics Concepts Inventory (Elrod, 2007). Items for economics were selected from items assessing knowledge of inflation from the Economic Model Questionnaire (Leiser & Briskman-Mazliah, 1996) and the Gambler's Fallacy Test (Donati, Chiesi, & Primi, 2015). The global geography questions were pulled directly from the compass direction task listed in Tversky (1981) and the Florida geography questions were constructed in a similar style (see Figure 1). The topics were split such that each discipline would have one topic tested at the first session and one topic tested at the second session. The number of questions in the pretests and posttests for across topics sets one and two were close to equal (30 or 33 questions).

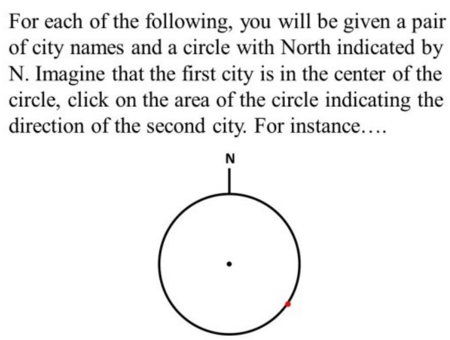

Figure 1: World and Florida Geography task and instructions. The task was identical for the World geography and Floridian geography tasks.

Epistemic Beliefs

Epistemic beliefs were measured using the Connotative Aspects of Epistemic Beliefs (CAEB) scale developed by Stahl and Bromme (2007). This inventory starts with an opening statement (e.g. "Knowledge in physics is…") and includes 24 adjective pairs (e.g. exact—vague) that could be used to describe knowledge in that given discipline on a scale of 1-7. For this scale, "1" (e.g. exact) aligns with a less advanced epistemic belief and "7" (e.g. vague) aligns with a more advanced epistemic belief. In accordance to the original scale, some items were reverse coded. Participants filled out one CAEB scale per discipline (physics, astronomy, genetics, economics, and geography, respectively). The score on the CAEB was the sum of responses, which could range from 24-168. There was one CAEB score per discipline from each participant and one overall CAEB score summed across disciplines.

Other Individual Differences

Reading ability was measured using the vocabulary component of Nelson-Denny (Brown, Fishco, & Hanna, 1993), a standardized reading comprehension ability test that has been used in prior research of similar nature (e.g. Ozuru et al., 2009). Participants were instructed to read the beginning of a sentence prompt and then choose one of five words that best completed the prompt (e.g. "Economic aid refers to…" (1) money, (2) information, (3) education, (4) farming, (5) culture). The final measure was the number of accurate responses on the test. Working memory was measured using an Automated OSPAN task (Unsworth, McAbee, Redick, & Hambrick, 2015). The OSPAN was electronically administered using E-prime software. To measure prior knowledge, participants were asked questions about which science and social science courses they took at the high school and undergraduate level and for how many semesters they took each course. Permission from the participants was also obtained to gather their ACT scores from university records. Fifty-nine participants in the final sample gave consent for their ACT scores to be obtained from university records. Science ACT scores were used in this investigation because this particular subtest of the ACT measures how students rapidly learn information from text and use it to answer multiple choice questions.

Procedure

Experimental Task

This study was a two-part, within-subjects design. Participants were informed in advance they would be participating in a two-session study with each session taking place approximately one week apart and lasting about one hour each. Time lag between sessions ranged from 5-9 days. During the first session, participants completed the CAEB scales, completed a pretest on one of the topics from each of the five disciplines, read text concerning the topics, completed an immediate posttest, and completed the Nelson-Denny. All tasks were completed on a Qualtrics survey that was saved on lab computers and links were not externally shared. To avoid carryover effects of text style, participants read either (1) expository or (2) refutation text during the first session.

The second session had a similar structure as the first. Participants answered the science background questionnaire, completed pretests, but on different topics from the first session, read text, completed an immediate posttest, and finished with the OSPAN task. For this second session, participants read text on different topics from the same five disciplines, but with a different text style they did not read during the first session (e.g. refutation if they read expository text in the first session). All tasks, except for OSPAN, were completed on a Qualtrics survey that was not available outside of the lab. OSPAN was operated with E-prime software. The order of topics tested was randomized across participants and order of Text Style was counterbalanced across participants.

Results

Description of model

The dependent measure for the study was an item-level coding of the accuracy of each item (1 = correct, 0 = incorrect). We conducted a mixed models logistic regression to predict the log odds of answering an item correctly. The model included participants and items as crossed random factors, with text style (expository/ refutation), test time (pretest/posttest), and epistemic beliefs (and all the two- and three-way interactions) as predictors. Binary predictors were contrast coded (-1, 1), and epistemic beliefs was centered around its mean prior to analysis. The regression models reported below contain random slopes of text style and test time on the participant level as well as the item level (models with the full complement of random slopes would not converge).

Main Model[1]

For the main model we analyzed the complete data set, collapsing across knowledge domains. Results revealed a significant effect of test time such that participants performed significantly better on the posttest compared to the pretest. This performance difference indicates that participants did in fact undergo conceptual change through the course of the experiment. Contrary to our hypothesis, however, neither text style nor epistemic beliefs were significant predictors of performance. None of the other predictors were significant.

Table 3: Analysis for the Main Model.

Fixed effects

Predictor	Coefficient	SE	t-value
Test Time	0.222	0.055	4.019**

Random slopes

Predictor	Variance	SD	χ^2
Text Style x ID	0.076	0.276	365.971**
Text Style x Item	0.027	0.165	144.49**
Test Time x Item	0.171	0.414	582.675**

Investigating Differences By Discipline

To further understand the item-level variability present in our data, we conducted exploratory analyses on each discipline. The effect of test time on conceptual change found in the main model was also noted in the Astronomy model (see Table 4), Economics model (see Table 5), and Geography model (see Table 6) such that participants performed significantly better on the posttest compared to the pretest, which indicates conceptual change within those content areas. These results, like the main model, however, cannot be attributed to text style.

There was also a small random effect of epistemic beliefs noted in the Geography model such that readers with less advanced epistemic beliefs exhibited more conceptual change than readers with more advanced epistemic beliefs. Interactions between text style and epistemic beliefs were also noted in the Astronomy, Economics, and Geography models such that refutation text and more advanced beliefs predicted higher amounts of conceptual change. As there are significantly fewer items available for analysis within each discipline model, the power is reduced and these findings should be treated as preliminary.

Contrary to what we found in the main model, there was also an interaction of test time and text style in the Physics model (see Table 7) and Genetics model (see Table 8) such that there were more accurate responses on the posttest and with expository text as opposed to pretest and with refutation text. This interaction effect within the models is rather small, thus we interpret it with caution.

Table 4: Analysis for Astronomy Model.

Fixed effects

Predictor	Coefficient	SE	t-value
Test Time	0.547	0.128	4.277**

Random effects and slopes

Predictor	Variance	SD	χ^2
Text Style x ID	0.156	0.395	220.20**
Test Time x Item	0.173	0.415	98.512**
Text Style x EB	<0.001	0.009	21.049$^+$

Table 5: Analysis for Economics Model.

Fixed effects

Predictor	Coefficient	SE	t-value
Test Time	0.215	0.053	4.101**

Random effects and slopes

Predictor	Variance	SD	χ^2
Text Style x ID	0.082	0.286	533.619**
Text Style x Item	0.022	0.148	137.797**
Test Time x Item	0.156	0.395	619.196**
Text Style x EB	<0.001	0.002	0.041$^+$

Table 6: Analysis for Geography Model.

Fixed effects

Predictor	Coefficient	SE	t-value
EB	-0.006	0.003	-2.017$^+$
Test Time	0.161	0.072	2.246$^+$
Text Style x Test Time	0.070	0.031	2.269$^+$

Random effects and slopes

Predictor	Variance	SD	χ^2
Text Style x ID	0.174	0.418	365.682**
Text Style x Item	0.020	0.141	34.919*
Test Time x Item	0.076	0.277	104.294**
Text Style x EB	0.0001	0.008	39.273*

[1] To conserve space, only significant results will be displayed in tables. Significance will be noted as follows: 0.05 level ($^+$), 0.01 level (*), 0.001 level (**). Epistemic beliefs will be abbreviated as EB. ID indicates variation of random slopes on the participant level. Random slopes will be discussed in a separate section.

Table 7: Analysis for Physics Model.

Fixed effects

Predictor	Coefficient	SE	t-value
Text Style x Test Time	-0.167	0.076	-2.203[+]

Random slopes

Predictor	Variance	SD	χ^2
Text Style x ID	1.219	1.104	690.845**
Test Time x Item	0.153	0.391	65.992**

Table 8: Analysis for Genetics Model.

Fixed effects

Predictor	Coefficient	SE	t-value
Text Style x Test Time	-0.167	0.076	-2.202[+]

Random slopes

Predictor	Variance	SD	χ^2
Text Style x ID	1.220	1.104	690.853**
Test Time x Item	0.153	0.391	65.993**

Exploration on the Item and Participant Levels

Random slopes included in the main model were significant predictors of response accuracy. Specifically, the random slope of text style across items and participants were significant, indicating that the difference between refutation and expository texts varied across participants and across test items. The significant random slope of test time across items also indicates that the pretest to posttest difference varied across items.

Random slopes for each discipline were also explored. For instance, across all disciplines, there was a significant random slope indicating a difference from pretest to posttest across items and a difference in refutation and expository text across participants. Also consistent with the main model, the Economics and Geography models showed a significant random slope of text style on the item level, indicating a difference in the refutation and expository text across items.

Despite these similarities in slopes across models, there were also some differences noted. For instance, there was a significant random slope of the interaction of text style by test time on the participant level observed in the Physics and Geography models, which indicates the differences in refutation and expository text interacted with differences in pretest and posttest responses across participants.

Overall, we see that the effects in each discipline model are not entirely uniform and could be attributed to a number of factors (e.g. certain items or domains may have been easier or more difficult for learners, and certain readers may learn better within the context of the experiment). These potential factors should be studied further. There were also significantly fewer items within each discipline model (ranging from 11-18 items) compared to the main model (containing 64 items), thus reducing the power available to make conclusions about the available data.

Preliminary Findings on Individual Differences[2]

Overall, participants with higher reading ability exhibited more conceptual change. We also note a similar relation in each discipline, with the exception of physics. We also found that readers with higher ACT scores exhibited more conceptual change in the astronomy and geography disciplines (see Table 9). Due to the small sample size (only fifty-nine participants granted researchers permission to access their ACT scores from university records), we interpret that relation with caution. There were no significant relations found between prior knowledge, working memory[3], and conceptual change.

Table 9: Individual Differences.

Discipline: Predictor	Coefficient	SE	z-value
Main: ND	1.076	0.197	5.455**
Astronomy: ND	1.428	0.393	3.630**
Astronomy: ACT	0.056	0.023	2.420*
Genetics: ND	1.111	0.404	2.750**
Economics: ND	2.179	0.714	3.051*
Geography: ND	0.854	0.293	2.917*
Geography: ACT	0.068	0.017	4.106**

Discussion

Our investigation of the effects of text style and epistemic beliefs on conceptual change builds upon previous studies by directly comparing how individual readers may learn differently after reading refutation versus expository text. In general, participants increased their scores between pre- and post-test, demonstrating that they did undergo conceptual change within the experiment. However, the effect of conceptual change did not vary across text styles (consistent with Hynd & Guzzetti, 1998) or readers' individual epistemic beliefs. Instead, random slopes reveal that there is variability in how individual students respond to different text types.

Analyses within disciplines revealed finer grained differences in this observed conceptual change. For instance, epistemic beliefs appeared to influence learning in astronomy, economics, and geography domains which is consistent with prior work that more advanced epistemic beliefs encourage learning (Hofer & Pintrich, 1997); however, these effects were not consistent across all domains. These findings suggest that epistemic beliefs are more important in some domains compared to others, and have implications for how educators may need to address epistemic factors in instruction. These differences in

[2] For the sake of brevity, only significant results are discussed. ND refers to Nelson-Denny. ACT refers to ACT science scores.

[3] Due to systematically missing data collected from the OSPAN, we were unable to reliably determine whether or not there was a relation between working memory and conceptual change.

epistemic beliefs across disciplines may also help account for disagreement in the role of this factor in prior work.

We next comment on why we failed to find differences between refutation and expository texts on conceptual change. Previous work on text style and conceptual change had suggested that not only the *style*, but also the *quality* of explanation matter for inducing conceptual change. Thus, it is possible that explaining why a misconception is false could induce conceptual change irrespective of the text style (refutation vs. expository; e.g. Lassonde et al., 2016). Future work will have to disentangle extent to which the text style and the explanation quality could induce conceptual change. Future work would also need to determine if the difference in text style and conceptual change can be reliably attributed to the tested disciplines. It is possible that certain concepts were simply easier to learn than others, and that some individuals were able to learn from the text more easily than others.

Preliminary investigation of individual differences reveals that reading ability and science ACT scores predict conceptual change in some disciplines. This finding is not too surprising considering that one's ability to learn from short science passages (as required on the ACT science section) and general reading ability would be generally useful in acquiring new information in this experiment. This result is also consistent with the idea that better readers are better comprehenders (e.g. Ozuru et al., 2009).

While the overall effect of conceptual change cannot be readily attributed to either text style or epistemic beliefs, we do observe some finer nuances of those two factors influencing conceptual change on the item and participant levels across disciplines. Thus, it appears that while individual participants do generally show more conceptual change on posttest performance, the style of text that induces the most conceptual change varies by reader, which would suggest that one text style may not be the best solution for all students.

Acknowledgements

This research is preregistered on Open Science Framework (https://osf.io/etpr7/).

References

Bailey, J., Johnson, B., Prather, E., & Slater, T. (2011). Development and validation of the star properties concept inventory. *International Journal of Science Education, 34*(14), 2257.

Brown, J., Fishco, V., & Hanna, G. (1993). *Nelson-Denny reading test: Manual for scoring and interpretation.* Chicago: Riverside Press.

Donati, M.A., Chiesi, F., & Primi, C. (2015). The role of probabilistic reasoning abilities on adolescent risk taking. *The Mathematics Enthusiast, 12*(1-3), 262-274.

Elrod, S. (2007). *Genetics concept inventory.* Retrieved from http://bioliteracy.colorado.edu/Readings/papersSubmittedPDF/Elrdo.pdf

Frick, R.W. (1998). A better stopping rule for conventional statistical tests. *Behavior Research Methods, Instruments, and Language, 25,* 43-58.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concepts inventory. *Physics Teacher, 30*(3), 141.

Hewson, P.W., & Hewson, M.G. (1984). The role of conceptual conflict in conceptual change and the design of science instruction. *Instructional Science, 31,* 1-13.

Hofer, B.K., & Pintrich, P.R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88-140.

Hynd, C., & Guzzetti, B.J. (1998). When knowledge contradicts intuition: Conceptual change. In C. Hynd (Ed.), *Learning from text across conceptual domains* (pp. 139-164). Mahwah, NJ: Lawrence Erlbaum Associates.

Kendeou, P., & O'Brien, E.J. (2014).The Knowledge Revision Components (KReC) Framework: Processes and mechanisms. Processing inaccurate information: theoretical and applied perspectives from cognitive science and the educational sciences, 353.

Kendeou, P., Walsh, E.K., Smith, E.R., & O'Brien, E.J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes, 51*(5-6), 374-397.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.

Lassonde, K.A., Kendeou, P., & O'Brien, E.J. (2016). Refutation texts: Overcoming psychology misconceptions that are resistant to change. *Scholarship of Teaching and Learning in Psychology, 2*(1), 67-78.

Leiser, D, & Briskman-Mazliach, R. (1996). The economic model questionnaire. Ben-Gurion University, Beer Sheva.

Posner, G.J., Strike, K.A., Hewson, P.W., & Gertzog, W.A. (1982). Accommodation of a scientific conception: Towards a theory of conceptual change. *Science Education, 67,* 489-508.

Stahl, E., & Bromme, R. (2007). The CAEB: An instrument for measuring connotative aspects of epistemological beliefs. *Learning and Instruction, 17,* 773-785.

Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education, 8,* 951-970.

Toplak, M.E., West, R.F., & Stanovich, K.E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking and Reasoning, 20*(2), 147-168.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology, 13,* 407-433.

Unsworth, N., Heitz, R.P., Schrock, J.C., & Engle, R.W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498-505.

van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology, 22,* 335-351.