Consciousness and Cognition 21 (2012) 1599–1616

Contents lists available at SciVerse ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Confidence in one's social beliefs: Implications for belief justification

Asher Koriat *, Shiri Adiv

Department of Psychology, University of Haifa, Israel

ARTICLE INFO

Article history:
Received 2 January 2012
Available online 18 September 2012

Keywords:
Belief justification
Subjective confidence
Self-consistency
Epistemology
Metacognition

ABSTRACT

Philosophers commonly define knowledge as justified true beliefs. A heated debate exists, however, about what makes a belief justified. In this article, we examine the question of belief justification from a psychological perspective, focusing on the subjective confidence in a belief that the person has just formed. Participants decided whether to accept or reject a proposition depicting a social belief, and indicated their confidence in their choice. The task was repeated six times, and choice latency was measured. The results were analyzed within a Self-Consistency Model (SCM) of subjective confidence. According to SCM, the decision to accept or reject a proposition is based on the on-line sampling of representations from a pool of representations associated with the proposition. Respondents behave like intuitive statisticians who infer the central tendency of a population based on a small sample. Confidence depends on the consistency with which the belief was supported across the sampled representations, and reflects the likelihood that a new sample will yield the same decision. The results supported the assumption of a commonly shared population of representations associated with each proposition. Based on this assumption, analyses of within-person consistency and cross-person consensus provided support for the model. As expected, choices that deviated from the person's own modal judgment or from the consensually held judgment took relatively longer to form and were associated with relatively lower confidence, presumably because they were based on non-representative samples. The results were discussed in relation to major epistemological theories – foundationalism, coherentism and reliabilism.

1. Introduction

There have been extensive discussions in the philosophy of mind on beliefs and their epistemic justification. Different positions have been advanced and there have been heated disputes over the strengths and weaknesses of each position. In this article, we address the question of belief justification from a psychological point of view. This examination will be carried out with reference to some of the current philosophical positions.

The present study can be seen to join the growing movement to investigate traditional issues in philosophy through empirical research. Indeed, philosophers subscribed to the emerging experimental philosophy approach (Bortolloti, 2008; Feltz, 2009; Loussouarn, Gabriel, & Proust, 2011) have begun to use methods of psychology, sociology, behavioral economics, and cognitive science to help shed some light on philosophically important questions (see Knobe, 2007; Michaelian, 2012; Thagard, 2002, 2009). Traditional philosophical questions have been tested empirically within a wide range of domains including moral responsibility and free will (Feltz & Cokely, 2009), ethics (Nichols, 2004), intentional action (Knobe, 2003), semantic categorization (Machery, Mallon, Nichols, & Stich, 2004) and epistemology (Feltz & Zarpentine, 2010; Swain,

* Corresponding author. Address: Department of Psychology, University of Haifa, Haifa 31905, Israel. Fax: +972 4 8249431.
E-mail address: akoriat@research.haifa.ac.il (A. Koriat).

1053-8100/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.concog.2012.08.008

Alexander, & Weinberg, 2008; Wright, 2010). In this study we focus on the question of epistemic justification, examining how empirical observations on people's convictions in their beliefs may bear on the cardinal philosophical approaches to belief justification. We hope that this project will contribute to the constructive interaction between empirical work on metacognition (Koriat, 2007) and philosophical work on epistemic justification.

Standard epistemological accounts of knowledge treat knowledge as a species of belief. In the philosophical Traditional Analysis of Knowledge (TAK), propositional knowledge is defined as Justified True Belief (JTB). This definition assumes three requirements. First, I cannot know *P* unless *P* is *true*. That is, truth is a defining property of knowledge. Second, I cannot know *P* if I do not even *believe* in *P*. Finally, my belief in *P* must be *justified*, that is, I must have some compelling account of the truth of *P*. This condition is intended to rule out mere lucky guesses (but see Gettier, 1963).

Philosophical theories of knowledge are concerned not only with descriptive theories of how people think but also in developing normative theories of how people ought to obtain knowledge (Thagard, 2009) as will be discussed later. Psychologists, who are more concerned with descriptive than normative issues, have been much more liberal in the use of the terms *knowledge*. This term has been sometimes used to refer only to a particular subjective feeling (e.g., Gardiner & Richardson-Klavehn, 2000), and sometimes to the correspondence between that feeling and the actual state of affairs (Kirkham, 1992). For example, some psychologists asked, "how do people know that they do not know?" (Glucksberg & McCloskey, 1981; Kolers & Palef, 1976). Also the question "how does one know that one knows" has been intended to mean "what are the psychological processes that lead to the feeling that one knows?" but it has been interpreted also to mean "why are people successful in knowing that they know?" Whereas the former meaning focuses on "knowing" as a subjective state irrespective of its correspondence to reality, the second meaning implies that what the person knows is true (see Koriat, 1993). The focus on "knowing" as a subjective state, has led experimental cognitive researchers to the examination of belief justification primarily in terms of the degree of confidence with which a belief is held, or in terms of the assessed probability that a proposition is correct.

1.1. Philosophical theories of belief justification

In philosophy, in contrast, there have been extensive discussions of what makes a belief justified or reasonable so that if true, it will be *known* to be so. In order to define justification we have to ask justification of what (Kirkham, 1992). In this article we focus on the epistemic justification of a belief in contrast to a moral justification and a utilitarian justification (see Reber & Unkelbach, 2010) According to *foundationalism*, some beliefs are justified on the basis of other beliefs. However, there are noninferential basic beliefs that are justified without owing their justification to any other beliefs (Van Cleve, 2005). Justified basic beliefs are generally assumed to include beliefs about simple logical or mathematical truths and beliefs about one's own mental state (however, moderate foundationalists include perceptional beliefs and memory beliefs as well). Thus, foundationalism assumes that there are justified basic beliefs, and that all justified nonbasic beliefs derive their justification from their relation to justified basic beliefs (Lemos, 2007).

By postulating the existence of basic beliefs, foundationalists escape the regress problem. According to this problem, any belief requires a justification, but the justification itself requires support, so that there is an infinite regress of (potential) justification (see Moser, 1988). The postulation of justified basic beliefs by foundationalism terminates the regress of justification.

An alternative solution of the regress problem has been proposed by proponents of the *coherentism* position. Coherentists challenge the linear conception of justification, which the regress argument presupposes. Instead, they hold a holistic conception of justification, in which a belief is justified by the way it fits together with the rest of the belief system of which it is a part (BonJour, 1985). Coherence theory avoids the problem of infinite regress without granting a special status to a particular class of beliefs.

A prominent feature of some of the foundationalist theories and the coherentism position is their *internalist* character: Epistemic justification is assumed to depend only on matters that are within the cognitive grasp of the believer. The assumption is that something that is not accessible to the believer's awareness cannot be used to justify a belief (see Chisholm, 1989; Conee & Feldman, 2001). This position has been challenged by proponents of *reliabilism* – one of the leading versions of the *externalist* view of epistemic justification. According to reliabilism, what makes a belief justified is the reliability of the causal process by which it was produced. That is, a belief is justified if the process that leads to its adoption has a high probability to produce true beliefs. A belief can be justified even if the process that leads to it and its reliability are inaccessible to the believer. What matters is reliability itself, and degrees of beliefs may vary depending on the degree of reliability (see Goldman, 1976,1988; Reber & Unkelbach, 2010).

In everyday life however, people often form their beliefs on the basis of imperfect informational sources. When the source is unreliable, as is often the case with testimony, metacognition is needed to solve the justification problem. Michaelian (2012) proposed to distinguish between the process in which some information is generated that can serve as the content a belief, and the process in which the accuracy of that content is evaluated in order to endorse it as a belief or reject it. It is the reliability of the latter, metacognitive process that counts in assessing process reliabilism. Assuming that people rely on a variety of criteria for evaluating the accuracy of the content of a potential belief, the endorsement of that belief can be conceptualized to depend on the overall subjective confidence in the accuracy of the information, so that when confidence surpasses a preset threshold, the belief can be endorsed. The implication is that the content of the information submitted to the endorsement mechanism may vary in a continuous manner along a dimension of subjective confidence. Indeed, philosophers

generally agree that justification is something that comes in degrees: One can have more or less justification for a belief (Engel, 1998; Feldman, 2003).

In this article, then, we focus on one's confidence in one's own belief, as it may provide insight into the psychological process of belief justification. We chose to focus on social beliefs because for statements describing social beliefs people do not usually have fixed ready-made responses and may change their response from one occasion to another. Between-individual and within-individual variation in responding is critical for testing some of the predictions from SCM.

In the empirical study to be reported, participants indicated their belief and then expressed their degree of confidence in that belief. Specifically, they were presented with statements depicting social beliefs, for example, "Old people are a heavy burden on society", "There is a supreme being controlling the universe". For each statement, they first made a *true*/*false* response according to whether they agreed or disagreed with the proposition. They then indicated their degree of confidence in their response on a 0–100 scale. The primary response – choosing *true* or *false* – reflects the person's belief. The secondary response – the assessed degree of confidence – is essentially a belief about a belief (Pelham, 1991), or a meta-belief (BonJour, 1985). As will be argued later, the assessment of the two types of belief rests on somewhat different cues. Our interest lies primarily in clarifying the metacognitive process by which people assess their certainty in their stated or constructed belief, which may provide some clues into the psychological process of belief justification. Therefore, the procedure we used allowed a separation between the cognitive process of specifying one's belief and the metacognitive process of assessing certainty in that belief. The task was administered to participants several times for reasons that will be explained later.

1.2. The metacognitive monitoring of one's own knowledge

The question of the basis of metacognitive judgments in general, and of confidence judgments, in particular, has been addressed in metacognition research primarily in connection with the monitoring of one's own knowledge during learning and remembering (Dunlosky & Metcalfe, 2009; Koriat, 2007). Three general approaches to the bases of confidence judgments have been distinguished: The direct-access approach, the information-based approach, and the experience-based approach (see Koriat & Levy-Sadot, 1999). The direct-access view assumes that confidence judgments are based on peoples' privileged access to the presence and strength of stored memory traces (see Schwartz, 1994). For example, when asked to decide whether a particular item was presented in a studied list, participants' confidence in their *yes/no* response is assumed to rest directly on the perceived familiarity of the item (see Van Zandt, 2000). In the information-based approach, metacognitive judgments are assumed to rest on an analytic process in which information retrieved from memory is consulted to reach an educated assessment of the likelihood that a proposition or a decision is correct. For example, in deciding between two possible answers to a question, confidence in a choice has been claimed to rest on the balance of evidence that is recruited in support of each of the answers (Griffin & Tversky, 1992; Koriat, Lichtenstein, & Fischhoff, 1980; McKenzie, 1997). This approach assumes that metacognitive judgments are inferential in nature rather than based on information that is read out directly from memory. The experience-based approach also assumes that subjective confidence is inferential in nature. However, the cues for confidence are seen to reside in the immediate feedback from task performance rather than in information retrieved from long-term memory. For example, confidence in the response to a *true*/*false* decision regarding a particular belief may be based on mnemonic cues that derive on-line from the experience of choosing the response (see Reber & Unkelbach, 2010; Sperber & Wilson, 1986). These cues may include the ease with which the choice was reached and the time it took to reach it (Robinson, Johnson, & Herndon, 1997; Zakay & Tuvia, 1998). Mnemonic cues are assumed to give rise automatically and unconsciously to a sheer metacognitive feeling (Kelley & Lindsay, 1993; Koriat & Ackerman, 2010; Koriat, Ma'ayan, & Nussinson, 2006) rather than to an educated judgment (see Proust, 2007, for a philosophical analysis).

In this study, we examine the processes underlying confidence in social beliefs within the Self-Consistency Model (SCM) of subjective confidence. This model was initially proposed for confidence in two-alternative forced-choice (2AFC) general knowledge questions (Koriat, 2012), but several results suggest that the model may apply to confidence in beliefs as well. According to the model, when participants have to choose the correct answer to a general-knowledge question (e.g., "Which country has more inhabitants (a) Italy, (b) France?"), they sample haphazardly various clues from memory and decide on the answer that receives the largest "support". The process is far from being analytic; rather, many of the clues that affect the choice consist of associations, hunches and images that are not readily expressed in the form of declarative propositions. Furthermore, how one determines pros versus cons often depends on a variety of heuristic cues such as fluency and familiarity. For example, studies of the illusory-truth effect indicate that the perceived truth value of a proposition is enhanced by the mere familiarity and fluency of a statement that are caused by a repetition of the proposition or by its context (Arkes, Hackett, & Boehm, 1989; Hasher, Goldstein, & Toppino, 1977; Unkelbach & Stahl, 2009).

According to SCM, the subjective confidence in one's decision is based on mnemonic cues that reflect the feedback from the process involved in attempting to reach a decision. These cues include the amount of deliberation and conflict that was experienced in reaching the decision, the amount of effort invested, and the speed with which the decision was reached. Collectively, these nonanalytic cues (see Jacoby & Brooks, 1984) capture self-consistency – the reliability with which the chosen answer was supported across the clues and considerations retrieved.

The effects of mnemonic cues on confidence are illustrated by the following observation: When participants were asked to list four reasons in support of their answer to a general-information question and then to indicate their confidence, they

reported *lower* confidence judgments than when they were asked to list only one supporting reason (Koriat, Nussinson, Bless, & Shaked, 2008). This result suggests that the effects of ease-of-retrieval – one reason is easier to retrieve than four reasons – can win over the effects of the declarative content of the reasons in affecting confidence judgments (see Schwarz, 2004).

1.3. Subjective confidence in social beliefs: the self-consistency model

We propose that SCM applies also to the process underlying confidence in social beliefs, although it might not apply to other types of beliefs such as determinate beliefs (see Koriat, in press-b). In the rest of the introduction, we describe the basic assumptions underlying the model. We then examine these assumptions in the light of the philosophical positions sketched earlier regarding belief justification. Finally, we describe a specific instantiation of SCM and derive the predictions that will be tested empirically.

When participants have to accept or reject a social belief, they usually do not retrieve a ready-made response. Rather, their decision depends on a constructive process, in which they search for evidence for or against the truth of the belief (Engel, 1998), drawing on information that they retrieve from within – thoughts, memories, and other beliefs. A similar view has been proposed with regard to people's evaluative attitudes toward a particular target: People construct their attitude on the spot, based on information that is most accessible at the time of judgment (Bless & Schwarz, 2010). Therefore, some degree of lability in the beliefs endorsed may be expected. According to SCM, this constructive process has much in common with that underlying statistical inference about the outside world. Participants are assumed to behave essentially like intuitive statisticians who attempt to reach a conclusion about a population on the basis of a small sample of observations drawn from that population. They sample several "representations" of the belief statement in turn, reach a tentative subdecision on the basis of each representation, and then make a final decision based on the predominant subdecision. The term "representation" is used broadly to refer to any consideration or interpretation that leans towards one of the two responses. Retrospective confidence in the response is based primarily on self-consistency – the reliability with which the chosen response is supported across the sampled representations. Thus, confidence increases with self-consistency in the same way that statistical level of confidence decreases with sample variance.

The subjective confidence in a belief is assumed to reflect the person's assessment of reproducibility – the likelihood that a new sample of representations will yield the same response. Subjective confidence can be said to incorporate a belief that is common among many researchers regarding the interpretation of statistical level of significance. A common misconception (see Schervish, 1996) is that in hypothesis testing, the correctness of the tested hypothesis, as well as the likely reproducibility of the observed result, is a monotonically increasing function of statistical level of significance.

Let us consider the metatheoretical assumptions underlying this proposal. First, the statistical (mis) conception underlying the interpretation of level of confidence embodies the logic that the internal consistency within a proximal sample of observations can serve as a cue for the extent to which a conclusion that is based on that sample corresponds to a property of the distal population. This logic escapes the regress problem. If a large proportion of the sampled observations tend to point to the same conclusion, we feel more confident that the conclusion is true. As discussed by Lewis (1946) and BonJour (1985, see Bovens & Hartmann, 2003), congruence or convergence can be seen as a form of coherence, so that coherence is essentially used as a cue for correspondence. Confidence judgments are construed subjectively as pertaining to correspondence – the likelihood that a proposition agrees with reality, but they are actually based on the coherence among the considerations that come to mind, in particular, the degree to which they converge in favoring supporting the proposition. A somewhat similar logic underlies coherentism theory in philosophy. Coherentism escapes the regress problem by assuming that a belief is justified by the system of other beliefs of which it is part.

SCM also implies the assumption underlying the reliabilism view, that beliefs can be justified when the process that leads to their adoption tends to produce true beliefs in a large proportion of cases (Goldman, 1986; Reber & Unkelbach, 2010). Reliabilism is compatible with the cue-utilization view in metacognition, which assumes that epistemic feelings such as subjective conviction or the feeling-of-knowing are based on the application of heuristics that have some degree of validity (Benjamin & Bjork, 1996; Koriat, 1997). Reber and Unkelbach (2010), in their analysis of epistemic justification, also emphasized the validity of processing fluency as a diagnostic cue for truth. When it comes to world knowledge, reliance on the self-consistency heuristic also tends to yield correct answers to 2AFC questions (Koriat, 2008). Furthermore, reliance on self-consistency yields confidence judgments that are diagnostic of accuracy. Consistent with reliabilism, it is assumed that reliance on self-consistency does not involve direct entailment, and operates bellow full consciousness in producing a sheer subjective feeling (Koriat, 2000).

Finally, self-consistency is output-bound (Koriat & Goldsmith, 1996): It is defined relative to the set of representations that are sampled at a particular time. As we note later, the sampling assumption underlying SCM implies that the representations that are accessed may vary across different occasions. Consequently, confidence is also expected to vary as a function of the internal consistency within the specific sample of representations underlying choice. The output-bound nature of self-consistency is somewhat consistent with the foundationalist position in that representations that are not sampled are not likely to affect the justification of the belief. It is unclear, however, whether internalists would consider heuristically-based confidence judgments as consistent with their position. Unlike what is implied by the foundationalism and coherentism positions as internalist views, the representations that are included in the sample are assumed to affect choice and confidence even if they are inaccessible to awareness.

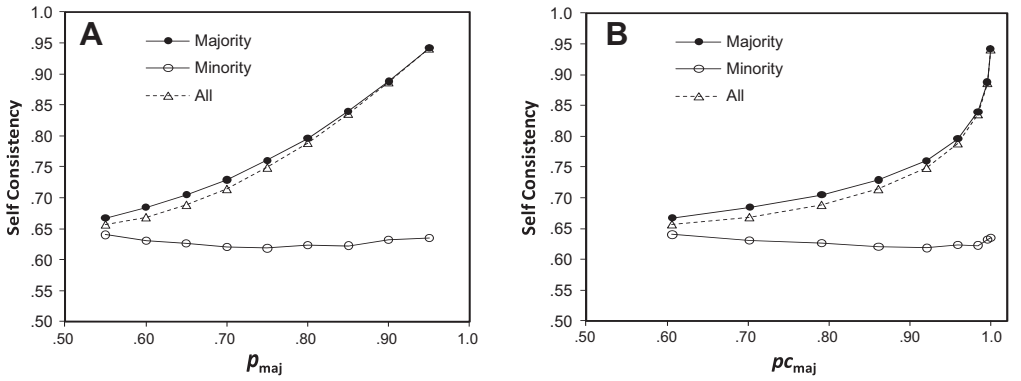

Fig. 1. Panel A presents self-consistency scores as a function of the probability of drawing a majority representation (P_{maj}). Panel B presents the same data as a function of the probability of choosing the majority option (PC_{maj}). Both Panels A and B are adapted from "The Construction of Attitudinal Judgments: Evidence from Attitude Certainty and Response Latency" by A. Koriat and S. Adiv, 2011, *Social Cognition*, *29*, pp. 586–587. Copyright © 2011 by Guilford Press. Adapted with permission.

1.4. The self-consistency model: predictions

Let us turn now to examine a specific instantiation of SCM, which will guide our investigation. SCM assumes that in making a response to a 2AFC item, representations are sampled and evaluated sequentially. The sampling continues until a preset sample size has been reached or until a series of draws yield the same subdecision a number of times in succession. In the specific SCM version to be examined (see Koriat & Adiv, 2011), we assume that the maximum number of representations (n_{max}) for each belief statement is 7, and that retrieval ends when 3 representations in a row favor the same option, in which case the outcome of the run-3 sequence determines the overt choice.[1] Confidence in the choice is assumed to depend on self-consistency, which is inversely related to the sample standard deviation. Assuming that p and q designate the proportion of subdecisions supporting the two response options, respectively, we use $1 - \sqrt{\hat{p}\hat{q}}$ as an index of self-consistency (range .5–1.0), calculated over the actual number of representations sampled (n_{act}). Response latency, in turn, is assumed to increase with actual sample size, n_{act}, that is, the number of representations drawn before an overt choice is made.

How can this model be tested? With regard to general-knowledge questions, Koriat (2012) proposed that in responding to 2AFC items, participants with the same experience tend to draw representations largely from the same, commonly shared population of representations associated with each item. Such is also true when the same person is presented with the same item on different occasions. If each representation favors one of the two answers, each item can be characterized by a probability distribution, with p_{maj} denoting the probability that a representation favoring the majority alternative will be sampled. Assuming a specific sample size (e.g., 7), p_{maj} for a given item may be estimated from pc_{maj}, the probability with which the majority alternative is chosen. This probability can be seen as a property of the item. It can be indexed operationally by the proportion of participants who choose the preferred alternative ("item consensus"), or by the proportion of times that the same participant chooses his/her preferred alternative across repeated presentations ("item consistency"). For example, for an item with a 30–70% between participant split of choices, item consensus will be 70%.

The results for general knowledge questions were consistent with the assumption of a consensually shared population of representations associated with each item. The present study is also based on the assumption that in making a *true/false* decision on a belief statement, participants draw a sample of representations from roughly the same commonly shared population of representations. Hence, p_{maj} for a given belief item may be estimated from the probability with which the majority response is chosen.

A simulation experiment was run (see Koriat & Adiv, 2011), which assumed a vector of binomial populations that differ in p_{maj}, with p_{maj} varying from .55 to .95, at .05 steps. For each population, samples of 7 representations each were drawn randomly, and the majority value in the sample was defined as the overt choice. However, when a sequence of 3 identical values occurred, the repeated value was defined as the overt choice. Responses were classified as "majority" when they corresponded to the majority in the population, and as "minority" when they corresponded to the minority value in the population. A self-consistency index and n_{act}, as described above, were calculated for each iteration.

The results of the simulation (Fig. 1A) bring to the fore the diagnostic value of the self-consistency index, which is assumed to underlie subjective confidence. Mean self-consistency ("All") increases with p_{maj}. More important, self-consistency is systematically higher for majority than for minority choices. This is so because for a sample that happens to favor a minority choice, the proportion of subdecisions favoring that choice will be smaller on average than when the sample favors the

[1] Using a stopping rule for data collection that depends on the outcome of previous observations is prohibited in orthodox hypothesis testing statistics (Dienes, 2011). However, to account for the inverse relationship between response time and confidence in several tasks, sequential-sampling models of human choice behavior have incorporated a stopping rule in which an overt choice is made when a criterion amount of evidence has been acquired (e.g., a criterion "run" of observations favoring a particular choice, Audley, 1960; see Baranski & Petrusic, 1998).

majority choice. For example, assume that 7 representations are drawn. For p_{maj} = .70, the likelihood that 6 or 7 representations will favor the majority choice is .329, whereas only in .004 of the samples will 6 or 7 representations favor the minority choice.

Fig. 1B, in turn, presents the same data but now as a function of pc_{maj}, the likelihood of choosing the majority option. Pc_{maj} is derived from the same simulation just mentioned. This figure specifies the predictions that can be tested when pc_{maj} is estimated either from item consensus or from item consistency. It should be noted that the simulation experiment yielded very similar functions for response speed as those depicted in Fig. 1 for confidence, when response speed was defined as the inverse of n_{act}.

Let us examine the predictions specified in Fig. 1, when pc_{maj} is estimated from item consistency – the proportion of times that a participant makes his/her most frequent choice across presentations of an item. Three features should be noted in this figure. First, confidence should increase monotonically with item consistency. Second, confidence should be higher on average for the frequent choice than for the rare choice. That is, when people fluctuate in their choice across presentations, they should express greater confidence when making their more frequent choice than when making their less frequent choice. Finally, confidence in the frequent choice should increase strongly with item consistency, whereas confidence in the rare choice should decrease but more shallowly so. Precisely the same pattern of results is expected for response speed.

The same pattern of results is expected when pc_{maj} is indexed by item consensus – the proportion of participants who choose the majority, consensual response on a particular occasion. This prediction is based on the assumption, noted earlier, that when people respond to an item, the representations that come to mind are sampled from a population that is largely commonly shared. The results obtained for general-information questions and perceptual judgments (Koriat, 2011, 2012) accord rather well with predictions that are based on this assumption. If this assumption also holds true with regard to social beliefs, we should expect confidence to be higher and response latency to be shorter when the person's choice is the consensual choice than when it is the nonconsensual, minority choice.

It should be stressed that these predictions follow from the assumption that participants' choice for each item depends on the majority option in *their* samples. The predictions are derived from a model that assumes random sampling. Hence, a higher confidence is expected for consensual choices independent of any social pressures that may be operating (Asch, 1955; Festinger, 1954).

Another prediction that will be tested is that confidence in a choice reflects the assessed likelihood that the same choice will be made if the same item is presented again. Assuming that this assessment is accurate, confidence in a choice in the first presentation of an item should predict the likelihood that the same choice will be made in subsequent presentations of that item.

As noted by Koriat (2012), SCM shares its basic assumptions with other models of confidence judgments (e.g., Gigerenzer, Hoffrage, & Kleinbölting, 1991; Juslin & Olsson, 1997; Stewart, Chater, & Brown, 2006; Vickers, 1970). However, the predictions derived from SCM follow uniquely from the conjunction of assumptions underlying this model. The basic prediction is that for any given item, confidence should differ systematically depending on which answer is chosen: When the majority alternative is chosen, it will be assigned higher confidence than when the minority alternative is chosen.

In our study, we examined these predictions using a *true/false* version of the Social Axioms Survey (SAS; Leung et al., 2002). This questionnaire has been found to yield inter-individual and inter-cultural differences, and norms are available for the Israeli population. The task was administered six times over 2 days, with self-report questionnaires interpolated between different administrations. In each administration, participants marked either *true* or *false* for each item, and indicated their confidence in their response.

2. Method

2.1. Participants

Participants were 41 Hebrew-speaking psychology undergraduates (26 females and 15 males) who were paid for their participation.

2.2. Stimulus materials

The experimental materials consisted of a 60-item Hebrew version of the Social Axioms Survey (SAS; Leung et al., 2002) that was prepared by Kurman and Ronen-Eilon (2004). The SAS measures general social beliefs, and was found to include five dimensions: Control by Fate, Reward for Application, Social Cynicism, Spirituality, and Social Complexity. Illustrative beliefs that are representative of these dimensions appear in Appendix A. The original response format of the questionnaire is a 5-point scale requiring a rating from "Strongly Disbelieve" to "Strongly Believe". This format was replaced by a 2-alternative, *true/false* format. In addition, after indicating their response, participants rated their confidence on a 0–100 scale (0-very unsure; 100-very sure).

Two self-report questionnaires were used as fillers between different administrations of the SAS questionnaire. These were the Need for Closure Scale (NFCS; Webster & Kruglanski, 1994) and the Rational–Experiential Inventory (REI; Pacini & Epstein, 1999).

2.3. Apparatus and procedure

The experiment was conducted individually on an IBM-compatible personal computer. It consisted of two sessions, one on each of two successive days. Each of the sessions included three blocks in which the entire set of 60 statements was presented. Participants were given the following instructions:

"We are conducting a study that concerns social beliefs, and will appreciate your cooperation. You will be presented with 60 statements. For each statement we would like you to decide whether it is true or false. There are no correct or wrong answers, so you should respond according to your own feelings. The 60 statements, all concerning social beliefs, will be presented one after the other. For each statement, we ask you to do the following:

Read the statement carefully, and press the left mouse key when you have finished reading it. At this time the response options ("True/False") will be added beneath the statement. You should indicate your response by clicking the chosen option according your own feeling, and then click the "confirm" box. Note that after pressing the "confirm" box you will not be able to change your response.

Immediately thereafter, you will be asked to indicate your confidence in the option you chose: After pressing the "confirm" box, a confidence scale (0–100) will be added beneath the alternatives. You are asked to mark your confidence in your response by sliding a pointer on a slider using the mouse. If you are very unsure in your response, mark 0. If you are very sure in your response, mark 100. Please try to make use of the full range of the confidence scale to express your confidence in your response. Press the "confirm" box after you mark your confidence. Note that after pressing the "confirm" box you will not be able to change your confidence rating. Do you have any questions before we begin?".

In each trial, the belief statement was presented on the screen until the participant pressed the left mouse key to indicate that he/she had finished reading it, at which time the response options ("True/False") were added beneath the statement. The participant indicated his/her response by clicking the chosen option. Response latency was measured, defined as the interval between the left-key press and the choice of a response. After pressing the "confirm" box, a confidence scale (0–100) was added beneath the alternatives, and participants marked their confidence by sliding a pointer on a slider using the mouse (a number in the range 0–100 corresponding to the location of the pointer on the slider was shown in a box). After clicking a second "confirm" box, the next trial began.

Following Block 1, the first part of the NFCS scale (21 items) was administered on the computer followed by Block 2. The second part of the NFCS scale was then presented, followed by Block 3. The procedure in the second session was identical to that of the first session except that the first part of the REI scale (20 items) was presented after Block 4, and the second part was presented after Block 5.

The order of the SAS items was determined randomly for each participant and block. In addition, each block was preceded by two warm-up items taken from a set of 12 items that were similar in content and format to those of the 60 SAS items. Each session of the experiment lasted about 45 min.

3. Results

Participants tended to make the same choice across the six blocks. Thus, if we focus on the choice made in Block 1, the likelihood of making that choice again over the next five blocks averaged .92 across all participants.

For most of the analyses to be reported, we first present the results for confidence and then those for response latency. Therefore, we begin by examining the relationship between these two dependent variables. As noted above, response latency was assumed to constitute a frugal behavioral indicator of self-consistency, and was expected to yield similar results to those of confidence judgments.

3.1. The relationship between confidence and response latency

In all of the analyses of response latency to be reported, latencies that were below or above 2.5 standard deviations (SD) from each participant's mean were eliminated (3.41%). We examined the relationship between confidence and response latency, focusing on the results from Block 1. All items were divided for each participant at the median of response latency in Block 1 (response latency for below-median and above-median responses averaged 1.97 s and 4.92 s, respectively, across participants). Confidence judgments averaged 83.77 for below-median response latencies and 72.36 for above-median response latencies, $t(40) = 10.03$, $p < .0001$. Thus, confidence was inversely related to response latency as has been found in previous studies (Kelley & Lindsay, 1993; Koriat, 2008; Koriat, Ma'ayan, & Nussinson, 2006; Robinson et al., 1997).

3.2. Confidence and latency as predictors of reproducibility

In this section, we examine the hypothesis that confidence predicts the reproducibility of the choice, that is, the likelihood that the same choice will be made on subsequent encounters with the same item. The confidence judgments in Block 1 were grouped into six categories and repetition proportion – the likelihood of making the same choice over the subsequent five blocks – was calculated across all participants. The results, pooled across participants and items, are presented in Fig. 2A. Indicated in this figure is also the number of observations in each category. It can be seen that the function is monotonic,

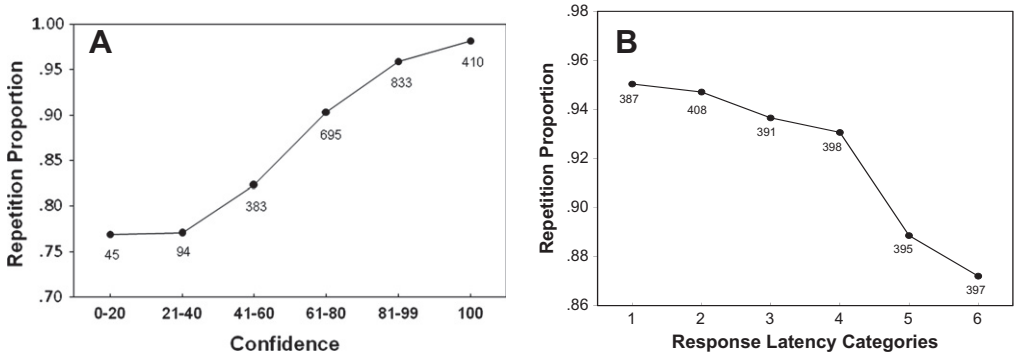

Fig. 2. Panel A presents the likelihood of repeating Block-1 responses as a function of confidence in these responses in that block. Indicated also is the number of observations in each confidence category. Panel B presents the same data for response latency.

indicating that repetition proportion increases with confidence in Block 1. The Spearman rank order correlation over the six values was 1.0, $p < .0001$.

To evaluate the reliability of the reproducibility results across participants, we divided the responses of each participant at the median of confidence judgments in Block 1. The percentage of response repetitions averaged 85.54% and 96.88% for below median and above-median confidence, respectively, $t(40) = 10.99$, $p < .0001$. The pattern of a higher repetition rate for above median than for below-median Block-1 confidence was exhibited by 38 of 40participants (1 tie), $p < .0001$, by a binomial test.

A similar analysis was carried out using latency as a predictor. The response latencies in Block 1 were grouped into six categories of about the same frequencies across participants. Fig. 2B presents repetition proportion for each of the response latency categories. The function is monotonic, indicating that response repetition decreases with response latency in Block 1. The rank-order correlation across the 6 points was -1.00, significantly different from 0, $p < .0001$.

In sum, both confidence and response latency are highly predictive of response repetition. This result is in line with the assumptions that confidence in a belief essentially monitors the likelihood that a response that was based on a given sample of representations is likely to survive if a new sample is drawn from the same population. Participants may fluctuate in the choice that they make on different occasions depending on the representations that are sampled on each occasion. However, because confidence depends on the reliability with which a choice is supported across the sample of representations, confidence is diagnostic of reproducibility: Choices that are associated with high self-consistency are more likely to be repeated than those that are associated with low self-consistency.

3.3. Confidence and latency as related to within-person response consistency

In this section, we examine the predictions of SCM for within-person analyses. These predictions derive from the assumption that in each block, participants base their choice and confidence on a small sample of representations drawn from the same population. Because the sampling is random, their choice may depart in some occasions from their most frequent choice. However, on these occasions confidence should be lower than on those occasions in which they endorse the frequent choices.

For each participant, the choices were classifies as ‘‘frequent’’ when they were made 4 or more times, and as ‘‘rare” if they were made twice or once across the six blocks. Fig. 3A presents mean confidence for the participant’s frequent and rare re-

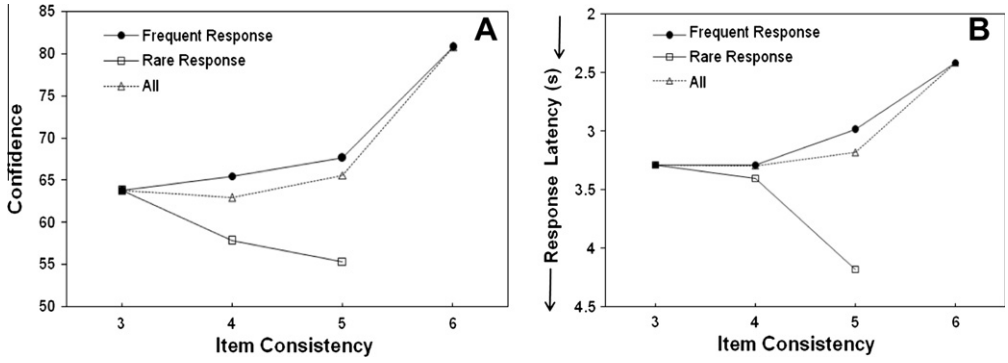

Fig. 3. Panel A presents mean confidence for the participant’s frequent and rare responses as a function of item consistency (the number of times that a response was made across all blocks). Panel B presents the same data for response latency.

sponses as a function of the number of times that the frequent response was chosen (“item consistency”). Included in this figure is also the mean for the tie items (item consistency = 3) and that of the “full consistency” items (item consistency = 6). As expected, mean confidence increased monotonically with item consistency. The results, however, differed for the frequent and rare responses. Focusing on item consistency 4 and 5, confidence was significantly higher for the frequent responses (66.45) than for the rare responses (56.59), $t(40) = 5.13$, $p < .0001$.

In addition, confidence in the frequent responses increased monotonically with item consistency, whereas confidence in the rare responses decreased with item consistency. This pattern is precisely what was predicted. Thus, participants were less confident when their response deviated from their modal response.

A similar pattern was observed for response latency (Fig. 3B). Focusing on item consistency 4 and 5, response latencies were shorter for the frequent responses (3.64 s) than for the rare responses (4.67 s), $t(40) = 2.81$, $p < .01$. In addition, for the frequent responses, response latency tended to decrease with item consistency, whereas for the rare responses it tended to increase with item consistency.

3.4. Confidence and latency as related to cross-person response consensus

As noted in the introduction, previous results with general-knowledge indicated that confidence is correlated with the consensuality of the answer (Koriat, 2008). In this section, we examine whether such is also the case for social beliefs. Using only the results from Block 1, we compared for each item the confidence for participants who chose the consensual response for that item with that of participants who chose the nonconsensual response.

For each of the 60 items, the response endorsed most often by participants in Block 1 was designated as the consensual response for that item. Item consensus was defined as the percentage of participants choosing the consensual response to that item. Item consensus averaged 77% (range 51–100%) across items. For one item, all participants gave the same response.

Fig. 4A presents mean confidence judgments for each of six item consensus categories (51–59%; 60–69%; 70–79%; 80–89%; 90–99%; 100%). Confidence judgments are plotted for consensual responses, for nonconsensual responses, and for all responses combined. Presented also is the number of items on which each mean is based. Mean overall confidence judgments tended to increase with item consensus: The correlation over all 60 items between mean confidence and mean item consensus was .58 ($p < .0001$).

However, when the consensual response was chosen it was endorsed with higher confidence than when the nonconsensual response was chosen: Across 59 items (as noted, for one item all participants chose the same response) confidence judgments averaged 78.90 for consensual responses compared with 68.68 for nonconsensual responses, $t(58) = 5.81$, $p < .0001$.

In addition, confidence in the consensual response tended to increase with item consensus whereas confidence in the nonconsensual response tended to decrease with item consensus. Indeed, across the 60 items the correlation between confidence and item consensus was .51 ($p < .0001$) for consensual responses, and −.35 ($p < .01$) for nonconsensual responses.

The possibility exists that the results just presented reflect a between-individual effect: Participants who tend to endorse the consensual beliefs tend also to be relatively more confident. However, when the confidence judgments were first standardized to neutralize chronic differences in confidence, the results were essentially the same as those obtained with the raw scores.

A similar analysis was carried out for response latency, and the results appear in Fig. 4B. Response latencies tended to decrease with item consensus: The correlation between mean response latency and mean item consensus across all items was −.46 ($p < .001$). Using only 58 items that had both consensual and nonconsensual responses, response latencies were longer for nonconsensual responses (4.11 s) than for consensual responses (3.29 s), $t(57) = 3.04$, $p < .005$. Across the 60 items, the correlation between latency and item consensus was −.29 for consensual responses, $p < .05$. The respective correlation for nonconsensual responses (across 58 items) was .04, *ns*.

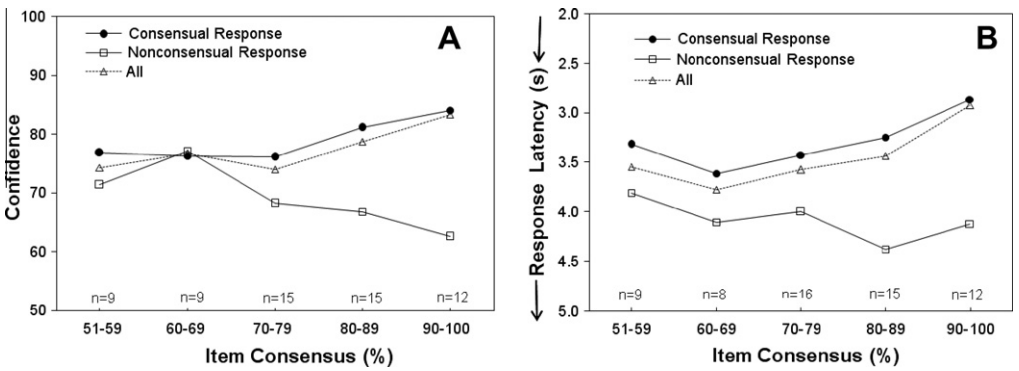

Fig. 4. Panel A presents mean confidence judgments in Block 1 for consensual and nonconsensual responses and for all responses combined as a function of item consensus (the percentage of participants who chose the majority response). Panel B presents the same data for response latency.

3.5. Confidence as a function of normative response consensus

In the previous section, the distinction between consensual and nonconsensual beliefs was based on the sample of participants who took part in this study. In this section, we examined whether the same results are obtained when the definition of consensuality is based on the Hebrew norms that are available for the SAS questionnaire. As indicated earlier, in the original version of the questionnaire, the response format is a 5-point Likert scale ranging from "1-Strongly Disbelieve" through "3-No Opinion" to "5-Strongly Believe". Using Hebrew norms that were obtained with this format (Kurman & Ronen-Eilon, 2004), the items were divided into five categories as follows: Strong *False* (mean less than 2.00), Moderate *False* (2.01–2.50), Neutral (2.51–3.50), Moderate *True* (3.51–4.00) and Strong *True* (above 4.0). For all items other than those in the Neutral category, participants' Block-1 responses were classified as consensual if the response agreed with the norm (e.g. *True* and the norm is above 3.51), and nonconsensual otherwise.

Fig. 5A plots mean confidence for consensual and nonconsensual responses for the Moderate Opinion (*True* or *False*) and Strong Opinion (*True* or *False*) categories as a function of item consensus. Included also is the mean for the Neutral category. The results revealed the expected difference between consensual and nonconsensual responses. Confidence in these responses averaged 81.82 and 66.82, respectively, across the "moderate opinion" and "strong opinion" items, $t(40) = 6.13$, $p < .0001$. Very similar results were obtained when the analysis was carried out on the standardized confidence scores.

The analysis was repeated using response latency as a dependent variable. The pattern was similar to that obtained for confidence judgments (see Fig. 5B). A comparison of mean response latency for consensual and nonconsensual responses (averaged 3.05 s and 4.17 s, respectively, across the "moderate opinion" and "strong opinion" items), yielded $t(40) = 4.31$, $p < .001$. Very similar results were obtained when the analysis was carried out on the standardized response latency scores.

In sum, when consensuality was defined in terms of the norms, the pattern of results obtained for both confidence and response latency accords with what was predicted by SCM. These results are consistent with the assumption that participants draw representations from a shared population of representations for each item, and that the characteristics of that population can be gauged from the norms.

3.6. Reliability of inter-item differences in choice and confidence

The assumption of SCM, that the representations associated with an item are commonly shared, implies that properties of items, notably, the likelihood of choosing the majority response and confidence in that response, are reliable across participants. Inter-participant reliability for Block 1was assessed using Cronbach's alpha coefficient (Crocker & Algina, 1986), which yielded a coefficient of .95 for response choice, and .71 for confidence judgments. These coefficients are remarkably high, supporting the assumption that participants base their choice and confidence on representations that are commonly shared.

SCM also assumes that participants draw representations from the same commonly shared population across repeated presentations of an item. Therefore, a correlation should be expected between within-person consistency and cross-person consensus: Responses that are consistently chosen by the same person should also be more likely to be chosen by others.

To examine this possibility, we calculated for each participant two scores for each item: (a) The proportion of times that the response made in Block 1 was repeated across the subsequent five blocks, and (b) the proportion of *other* participants (out of 40) who made that response in Block 1. These two scores were then averaged for each item across participants. The correlation between the two averages was .55 across the 60 items, $p < .0001$. Furthermore, the confidence of a participant in the choice made in Block 1 predicted the likelihood that *other* participants will make the same choice: The correlation was .58, $p < .0001$. These high correlations suggest that indeed consistency and consensus disclose roughly the same parameter that is associated with an item, a parameter that is relevant to confidence.

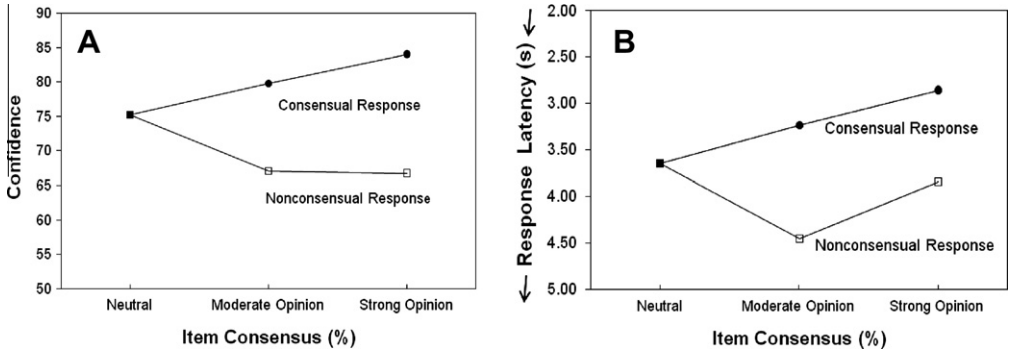

Fig. 5. Panel A: Mean confidence in Block 1 for consensual and nonconsensual responses defined in terms of the Hebrew norms as function of item consensus (the percentage of participants who chose the majority response). Panel B presents the same data for response latency.

McCloskey and Glucksberg (1978) also observed a correlation between cross-person consensus and within-person consistency in a task that required participants to say which of a list of items belonged in each of several semantic categories. It was proposed that deciding on category membership is an inherently probabilistic process (Hampton, 2011).

3.7. The joint effects of within-person consistency and between-person consensus

Given that confidence is associated with both consistency and consensus, it is of interest to examine the joint contribution of the two variables to confidence judgments. For each participant, the response to an item in Block 1 was classified as (a) consensual or nonconsensual based on the responses made to that item in Block 1 by the remaining 40 participants, and (b) frequent or rare, depending on its within-participant frequency across Blocks 2–6. A Consensus X Consistency ANOVA on mean confidence judgments was conducted, based only on 30 participants who had all 4 means. The ANOVA yielded $F(1,29) = 15.61$, $MSE = 73.27$, $p < .001$, for consensus, $F(1,29) = 64.73$,$MSE = 126.65$, $p < .0001$, for consistency, and $F < 1$, $MSE = 103.60$, *ns*, for the interaction. Although confidence increased with both factors, the overall effect of consistency was considerably stronger, amounting to 16.53 points, in comparison with 6.17 points for the effects of consensus.

These results suggest that an index of pc_{maj} that is based on within-person consistency in beliefs provides better clues to the dynamics of confidence judgments than an index that is based on cross-person consensus. It is interesting to note that in a similar analysis carried out on confidence in perceptual judgments (Koriat, 2011), the contribution of item consensus to confidence was about the same as that of item consistency. In contrast, the results for social attitudes (Koriat & Adiv, 2011) and for personal preferences (Koriat, in press-a) yielded a pattern similar to that observed for social beliefs. Presumably, in the case of beliefs and attitudes, within-person consistency is a better diagnostic of the self-consistency underlying choice and confidence than is cross-person consensus.

4. Discussion

In this study, we examined the psychological processes underlying the subjective confidence in one's beliefs. Confidence judgments represent meta-beliefs. However, they are assumed to capture the processes underlying the construction of the object belief itself. It was proposed that when people are asked to indicate their social beliefs, they do not usually retrieve a ready-made response, but deliberate before responding. They tend to construct their belief on the spot on the basis of the most accessible information. A similar view has been proposed with regard to people's positive or negative social attitudes towards a particular object (e.g., capital punishment). It has been argued that attitudinal judgments are constructed on the basis of the information accessible when making the judgment rather than being read out directly from memory (Bless & Schwarz, 2010; Schwarz, 2007; Schwarz & Strack, 1985; Wilson & Hodges, 1992). Indeed in a study by Koriat and Adiv (2011) certainty in one's attitudinal judgment was found to be sensitive to both the stable and variable aspects of attitudinal judgments, providing a clue to the on-line construction of these judgments.

Once a belief has been constructed, confidence in the belief was assumed to rely primarily on mnemonic cues that reflect self-consistency – the consistency with which the belief is supported by the representations and considerations that come to mind. The sampling assumption underlying SCM implies some fluctuation in the belief that is constructed on different occasions but also some commonality across these occasions. The results suggest that confidence judgments are sensitive to the properties of the sample of representations that are *accessible* in making a choice but also to the properties of the population of representations that are *available* in memory in connection with a belief.

4.1. Confidence in social beliefs: evidence for SCM

SCM attempts to model the retrospective portrayal of the process of belief construction. This portrayal is assumed to reflect primarily the degree of support for the chosen option in comparison with the alternative option. The version of SCM that was tested in this article is clearly an oversimplification. It assumes a random sampling of a maximum number of representations for each item, each representation is assumed to deliver a binary subdecision, and each subdecision is assumed to make an equal contribution to the self-consistency index that is assumed to underlie subjective confidence. The model could be modified in several ways to capture some of the complexities of the underlying process. However, the model in the form that it was presented made the derivation of predictions quite straightforward, and somewhat surprisingly, the results confirmed rather well the qualitative pattern that was predicted.

An essential component of the model is the assumption that participants draw a sample of representations of each item from a population of representations that is largely commonly shared. Furthermore, it was assumed that on repeated presentations of an item, participants draw their samples from the same population. These assumptions may seem reasonable in the case of general knowledge questions, but it is impressive that they are also largely true of social beliefs. Thus, the choice of a belief and the confidence in the belief were quite reliable across participants. In addition, beliefs that were endorsed more consistently and more confidently by the same person across blocks were more likely to be endorsed by other participants than those that were endorsed less consistently and less confidently. These results suggest a commonly shared core of representations associated with each item.

Confidence and response speed in Block 1 were found to be highly predictive of the likelihood of making the same choice in subsequent blocks (Fig. 2). This result accords with the assumption that confidence judgments represent an assessment of the likelihood of making the same choice when the item is presented again.

As far as the within-person analyses are concerned, mean confidence increased with increasing within-person consistency. In addition, as predicted by SCM, the more frequent beliefs were endorsed with stronger confidence than the less frequent beliefs. A similar pattern was observed for response speed (Fig. 3).

The results for cross-person consensus yielded a pattern that was qualitatively similar to that of response consistency. Mean confidence increased with cross-person consensus, and consensual beliefs were associated with stronger confidence than nonconsensual beliefs (Fig. 4). This pattern was replicated when consensuality was defined on the basis of the norms rather than on the basis of the results of the experiment itself (Fig. 5). Once again, the results for response latency yielded a very similar pattern to that observed for confidence.

Overall, the results are consistent with the prediction of SCM (Fig. 1) that confidence judgments should be lower when a choice departs from what follows from the population of representations associated with an item. The majority choice was inferred from the distribution of choices either across participants or within participants. Interestingly, the effects of within-person consistency on confidence were stronger than were those of cross-person consistency. This pattern suggests systematic inter-individual differences in the population of representations associated with each social-belief item so that within-person consistency is a better index of self-consistency than is cross-person consensus.

Let us now examine the metatheoretical assumptions underlying these results and their interpretation in terms of SCM.

4.2. Sampling information from within

The distinction in epistemology between rationalism and empiricism implies a separation between two possible sources of knowledge – reason and sense experience (Carruthers, 1992; Markie, 2008). The view advanced in this article, however, assumes that the process underlying confidence is similar for beliefs derived from the two sources (see Koriat, in press-b). Thus, confidence in a social belief (e.g., "Old people are usually stubborn and biased") has much in common with confidence in a proposition that can be validated on the basis of sense experience (e.g., which of two lines is longer; Koriat, 2011). In attempting to validate one's answer to a question about a social belief, a general-knowledge question, an episodic memory question, or a perceptual comparison task (in the absence of access to various external aids) one must make do with a variety of pieces of information accessed from within (Wright, 2010). Consistent with this assumption, the pattern of results obtained in this study for confidence in social beliefs was very similar to that observed for perceptual judgments. There is no question that different types of considerations are explored in reaching a decision in different domains such as world knowledge, social attitudes and perceptual judgments (Koriat, 2012). However, in all of these domains, the judgment is based on reasoning, that is, on a mental rather than a physical exercise. The similarity of the results across these domains also supports the assumption that confidence rests on cues that are largely indifferent to the specific content of the considerations that are examined in reaching a decision.

4.3. Self-consistency and reproducibility

Although the validation of one's knowledge and beliefs is based on retrieving information from within, SCM assumes that the underlying psychological process has much in common with that of making inference about the outside world. Specifically, it shares certain similarities with the statistical procedure that is used to test a hypothesis about a distal population based on a sample of observations drawn randomly from that population. The view of belief construction and justification as dependent on random sampling of clues from within is perhaps alien to most philosophical views of belief justification. In all likelihood, the process is much less random than postulated by SCM (see Koriat, 2012). Nevertheless, the results confirmed the gross predictions made by SCM. The assumption is that because of the limitations of the cognitive system, only a small number of representations are considered when constructing a belief, and the retrieval of these representations is subject to a certain degree of randomness.

The sampling assumption underlying SCM also implies that although beliefs aim at truth (Engel, 1998), and confidence is construed subjectively as pertaining to the truth of a belief, the subjective confidence in a belief actually monitors its reproducibility. It monitors the likelihood that the same choice will be made in the future. The results support the accuracy of confidence and response latency in predicting the likelihood of repeating the same choice in subsequent presentations of the item. In a sense, reproducibility is an operational definition of truth when it comes to inference about a population based on a sample of observations.

4.4. The relationship to philosophical theories of the justification of beliefs

Let us now examine the view advanced in this article in the light of the philosophical theories mentioned in the introduction regarding the justification of beliefs. To the extent that confidence in social beliefs provides some clues into the psychological justification of these beliefs, how does the view presented in this article fare with the philosophical views about belief justification?

The present study suggests that the logic underlying belief justification differs markedly from that postulated in philosophical discussions. In particular, subjective confidence in a belief appears to rely much less heavily on logical and inferential entailment than has been assumed by foundationalists. Indeed, previous studies suggest that the perceived truth of a statement is influenced by manipulations that do not concern the *content* of the belief. Thus, truth judgments are enhanced by manipulations that increase the mere familiarity and fluency of a statement (Bacon, 1979; Dechêne, Stahl, Hansen, & Wänke, 2010; Hasher et al., 1977; Unkelbach & Stahl, 2009). Statements of the very same content are judged as more probably true when they are written in concrete language than when they are written in abstract language, possibly because of the greater perceived vividness of concrete statements (Hansen & Wänke, 2010). Confidence is also affected by fluency. Thus, confidence in the answer is higher when questions are printed in a font that is easy to read than when printed in a font that is difficult to read (Alter, Oppenheimer, Epley, & Eyre, 2007). Imagining a childhood event enhances confidence that the event had occurred (Garry, Manning, Loftus, & Sherman, 1996), and confidence in the answers increases with the speed of retrieving or selecting these answers (e.g., Kelley & Lindsay, 1993; Koriat, 2008; Koriat, Ma'ayan, & Nussinson, 2006; Robinson et al., 1997).

The view presented in this article is consistent with these findings in emphasizing the idea that confidence judgments rest heavily on mnemonic cues that stem from the process of belief construction. The construction of a belief generally involves the retrieval of various considerations that may have a propositional form. However, retrospective confidence in that belief is assumed to depend not so much on the content of these considerations but on the overall impression that is formed about self-consistency. Not only is self-consistency a structural, contentless cue that does not involve the type of entailment relations emphasized by internalist philosophers, but also reliance on self-consistency does not involve an explicit inference. Rather, self-consistency and response speed are assumed to instill directly a feeling of certainty or doubt.

The psychological work in metacognition, however, is most consistent with reliabilisim as an externalist position in philosophy (see Proust, 2008). The assumption that metacognitive judgments rest on a variety of mnemonic cues such as fluency, familiarity, and accessibility, implies reliance on general-purpose heuristics rather than on domain-specific propositions (Koriat et al., 2008).

Let us point out some specific similarities between the self-consistency model and reliabilism. First, the proposal that confidence in a belief rests on self-consistency implies that a belief might be justified because the general causal process that led to its adoption is cognitively reliable, that is, produces true beliefs in a high proportion of cases. Reber and Unkelbach (2010) also endorsed the same assumption in discussing the role of processing fluency. They asked whether reliance on processing fluency as a cue to truth is epistemically justified. In their review they noted that on the one hand, manipulations that enhance the processing fluency of a statement increase the likelihood that it would be judged as true. On the other hand, a positive correlation exists in the real world between fluency and truth, so that processing fluency has some validity as a cue for truth. These observations were taken to imply that in terms of Goldman's (1986) view of reliabilism, processing fluency can be said to provide above-chance probability of true beliefs. In the same manner, the degree of consistency with which a statement is supported across different representations is generally diagnostic of the likelihood that the statement is correct. Thus, self-consistency can be said to provide epistemic justification as a reliable cue for truth.

Second, for reliabilism, what matters is reliability itself and not any awareness on the part of the person that the process is reliable (Goldman, 1979). Consistent with this view is the assumption among metacognitive researchers that mnemonic cues affect metacognitive judgments automatically and bellow full consciousness (Koriat, 2000). Indeed, in line with the externalist emphasis of reliabilism, reliance on mnemonic cues as a basis for confidence has been observed even for young (2nd-grade) children, who possibly lack the kind of sophistication that is needed for reasoned inferences (Koriat & Ackerman, 2010).

Reliable belief-producing processes, however, are not guaranteed to be truth conducive. This is also true of mnemonic-based judgments (see Michaelian, 2012; Reber & Unkelbach, 2010). When the ecological validity of a mnemonic cue is less than perfect, reliance on that cue is bound to yield incorrect decisions in some cases. Metacognition research has documented many cases in which reliance on mnemonic cues yields incorrect metacognitive judgments (e.g., Benjamin, Bjork, & Schwartz, 1998; Brewer & Sampaio, 2012; Busey, Tunnicliff, Loftus, & Loftus, 2000; Chandler, 1994). This is true of self-consistency as well, which may lead to false beliefs in some cases (Koriat, 2011, 2012).

4.5. Coherence and correspondence

Goldman (1986) argued that although coherence is not a required criterion for the justification of beliefs, it is a derivative standard. If a belief is true, it is required that it is consistent with other true beliefs. Thus, reliabilism implies the requirement of coherence of beliefs (Reber & Unkelbach, 2010). Self-consistency can be seen as a form of coherence. Hence, SCM implies that coherence is used as a cue for correspondence. Let us examine this idea with regard to statistical hypothesis testing.

The idea that coherence can serve as a cue for correspondence underlies the oft-criticized interpretation of statistical level of confidence (Schervish, 1996; see Dienes, 2011). A common conception among researchers is that the level of significance with which a null hypothesis is rejected (as indexed, for example by Fisher's p value) represents an assessment of (a) the probability that the alternative hypothesis is "true", and (b) the likelihood of replicating the effect that was obtained. Truth, of course, refers to the extent to which a conclusion that is based on a sample of observations applies to the population as a whole and hence is replicable. Thus, correspondence and replicability are intimately tied.

The calculation of statistical level of confidence can be seen to represent a specific theory of belief justification that ties together coherence and correspondence. Coherence in this case is narrowly defined in terms of internal consistency – the degree of agreement among the randomly sampled observations in supporting a specific conclusion. Thus, degree of coherence within a proximal sample is used as a cue for the likelihood that the conclusion that is based on that sample is true of the distal population.

In philosophical discussions, there has been a heated debate whether coherence can be said to be truth conducive, and whether coherence in the sense just described can be said to represent a *coherentism* theory of justification (e.g., Olsson, 2001, 2002). In discussing this theory, BonJour (1985) referred to the example mentioned by Lewis (1946) of several relatively unreliable witnesses who report the same story. According to Lewis, the congruence of the reports lends high credibility to these reports even if the credibility of each single report is very low. BonJour went further to argue that agreement between witnesses can be taken to justify the hypothesis of truth telling even if the reports are more likely to be false than true, provided that these reports can be assumed to be independent of each other. This argument has been criticized, claiming that coherence in itself does not enhance credibility if the individual reports are unreliable (Huemer, 1997; Olsson, 2002), and that coherence cannot generate credibility from scratch but can only amplify the justification of a belief that is already independently justified on some other grounds.

These and other reservations notwithstanding, the epistemology underlying statistical hypothesis testing incorporates the conception that the credibility of a conclusion about a population increases as a function of the degree to which it is supported across a number of independently sampled observations. Convergence is a holistic property of the collective sample of observations, and is assumed to enhance the probability that a hypothesis is correct even if each observation alone does not provide sufficient justification of the belief. The convergence conception of coherence has played an important role in Bayesian approaches to coherence (see Bovens & Hartmann, 2003; Bovens & Olsson, 2000). However, we dwelled here on the so-called orthodox statistical procedure rather than the Bayesian approach (see Dienes, 2011) because it seems to capture better the process assumed to underlie subjective confidence in a 2AFC decision.

Although there is some similarity between self consistency and the notion of coherence in coherentism theory of justification, there are important differences that should be noted. First, self-consistency pertains to a much more local level than that implied by coherentism theory. Specifically, it pertains to the coherence between the sampled representations underlying a constructed belief rather than to the coherence of that belief with other beliefs. Second, the notion of coherence assumed by SCM is much looser than that implied by coherentists, because what is activated during the construction of a belief is generally an assortment of images, memories, beliefs, associations, and thoughts that cannot always be expressed in a propositional form. Therefore, confidence may rest on nondoxastic sources, unlike what is assumed by coherentism, as an internalist theory of justification (see Davidson, 1986). For example, some of the factors discussed by Gendler (2008) under the concept of alief may also influence one's confidence in a belief even if they do not affect the stated belief itself.

Finally, SCM assumes that what matters is only the coherence within the set of beliefs and thoughts that are activated during the attempt to construct a belief. In this respect, coherence can be said to be output-bound (Koriat & Goldsmith, 1996): It is relative to the set of clues that are accessible. This view departs from BonJour's (1985) proposal that the coherence of beliefs at a particular moment is not a sufficient condition for their being justified; the system of beliefs must be coherent in the long run.

In sum, because people have no access to the object of their beliefs over and above what they believe about it, they rely on a fast assessment of overall coherence (see Bolte & Goschke, 2005) as a basis for their judgments about correspondence. In terms of Polanyi's (1958) terminology, the "object" of metacognitive judgments is correspondence, but the "tool" is coherence. Reliance on coherence as a cue for correspondence escapes the regress problem of belief justification without postulating basic beliefs.

4.6. Consensus and the consensuality principle

In our study, both coherence (self-consistency) and confidence were found to be related to inter-person consensus. Furthermore, consensus was also generally related to accuracy in domains for which the answers have a truth-value (Koriat, 2011, 2012). These observations may seem pertinent to discussions in philosophy in which inter-person agreement was treated as a form of coherence that boosts confidence (BonJour, 1985; Lewis, 1946). They seem also to be pertinent to the issue whether coherence is truth conducive (Huemer, 1997; Klein & Warfield, 1994; Olsson, 2005). However, the theoretical status of inter-person consensus in SCM is very different from what is implied by these philosophical discussions, as will now be clarified.

Let us first examine the question of accuracy (or "truth") of one's choices, which of course, could not be evaluated for the social and metaphysical beliefs used in this study. SCM was originally developed to explain the accuracy of confidence judgments. In studies of confidence in general knowledge, a positive within-person confidence–accuracy correlation is typically observed, suggesting that participants are successful in monitoring the correctness of their answers. Koriat (2008), however, reported results indicating that the confidence–accuracy correlation is a consequence of the fact that in 2AFC general-knowledge tasks, participants' choices are more likely to be correct than wrong. However, when a relatively large set of consensually wrong items is used, for which most participants choose the wrong answer, the confidence–accuracy correlation is actually *negative*: Participants are more confident when they choose the wrong answer than when they choose correct answer. Thus, confidence judgments are correlated with the consensuality of the answer – the proportion of other people who

chose that answer – rather than with its correctness. This generalization, which was labeled the *consensuality principle*, has been confirmed across several different domains (see Koriat, 2008, 2011, 2012).

What then is the status of inter-person consensus? In many tasks that measure general knowledge, memory or perception, people are more correct than wrong. For such tasks, the consensual answer is the correct answer, and hence participants' confidence is diagnostic of the correctness of the answer. Furthermore, the wisdom-of-crowds phenomenon suggests that information that is aggregated across participants is generally closer to the truth than the information provided by each individual participant (Galton, 1907; Surowiecki, 2005). Thus, inter-participant consensus is generally diagnostic of correctness, and is also correlated with high confidence.

However, consensus is conceptualized by SCM only as a correlate of self-consistency. As noted earlier, SCM explains the confidence–consensuality relationship by assuming that cross-person consensus is diagnostic of the internal consistency within the sample of representations underlying choice and confidence. This view differs from the suggestions in philosophical discussions that inter-participant agreement, as such, should boost confidence in a belief (BonJour, 1985). It also departs from suggestions by social psychologists that stressed the *causal* role played by social consensus in supporting confidence. Indeed, several findings indicated that people feel more confident in their views when they learn that others share these views (e.g., Clarkson, Tormala, DeSensi, & Wheeler, 2009; Luus & Wells, 1994; Orive, 1988). These findings are in line with social comparison theory, which assumes that in the absence of objective criteria, people assess the correctness of their views by comparing them to those of others (Festinger, 1950). Fazio (1979) also proposed that social consensus is taken to reflect validity. In addition, Bassili (2003), who examined the time it took participants to indicate their response to items that measure social opinions, found that participants who hold a minority opinion tend to express that opinion less quickly than people who hold the majority opinion. He attributed this effect to social inhibition in expressing opinions that depart from those of most others.

According to SCM, in contrast, consensual beliefs should be expressed more quickly and should be held with stronger confidence regardless of any social pressure towards conformity. This is not to say that social pressure does not affect response latency and confidence. Rather, we have shown that a model that assumes random sampling of representations from a commonly shared pool of representations is bound to yield stronger confidence in consensual than in nonconsensual beliefs (Fig. 1). This difference should occur irrespective of the *content* of these beliefs. It is important to stress that this difference is also expected to occur irrespective of the *correctness* of the belief (when a criterion of accuracy is available), as clearly demonstrated by the results testifying for the consensuality principle (Koriat, 2012).

Turning finally to the question of accuracy, to what extent can self-consistency be said to be truth conducive? As noted in connection with statistical hypothesis testing, many researchers behave as if a hypothesis is more credible the higher the level of significance with which the null hypothesis can be rejected. The implication is that given a particular sample size, confidence in an experimental effect increases as a function of the consistency with which the effect was supported across observations. Note, however, that a critical element is the assumption of *independence* among observations, as also stressed in philosophical discussions (BonJour, 1985; Lewis, 1946; Olsson, 2002). Of course, the assumption of independence is less tenable when information is sampled from within. The fact that whenever one thinks of the capital of Australia, "Sydney" comes consistently and persistently to mind, does not guarantee that Sydney is the correct answer. Self-consistency is not *inherently* diagnostic of accuracy, as demonstrated by the results on the consensuality principle, so that reliance on self-consistency is liable to breed unwarranted certainty in many cases. Indeed, Koriat (2012) argued that the overconfidence bias that has been observed in studies of confidence in general knowledge (Lichtenstein, Fischhoff, & Phillips, 1982) stems largely from the fact that participants rely on self-consistency as a cue for validity. Self-consistency is practically always higher than validity.

In this article, we focused on social beliefs. A question that remains open is whether SCM can also be applied to confidence in *a priori* beliefs. For such beliefs, there is little variance between the outcomes of different representations and perhaps this is the reason for the strong conviction with which these beliefs are held. However, confidence in such beliefs might be based on a qualitatively different basis. For example, BonJour (1985) claimed that coherence is not necessary for the justification of a priori beliefs.

In conclusion, the present study examined the processes underlying the subjective confidence in one's social beliefs under the assumption that confidence can shed some light on the process of belief justification. The results were examined in the light of the dominant philosophical theories of belief justification. In recent years, there has been a great deal of mutual influence between philosophers and empirical cognitive psychologists (Thagard, 2002). Although philosophers discussed justified beliefs from a normative perspective, their discussions rely heavily on (what they assume to be) people's intuitions about what can be considered as "justified beliefs" or "knowledge". It is hoped that the work presented will promote the ongoing constructive influence between cognitive psychology and philosophy.

Acknowledgments

We are grateful to Jenny Kurman for providing us with the Hebrew version and the Hebrew norms of the Social Axioms Survey. We thank Dana Klein and Rinat Gil for their help in the analyses, and Hila Sorka for conducting the experiments. Support for this project was received from the Max Wertheimer Minerva Center for Cognitive Processes and Human Performance, University of Haifa.

Appendix A. Illustrative items used in the study

1. Young people are impulsive and unreliable.
2. One's behaviors may be contrary to his or her true feelings.
3. Every problem has a solution.
4. Belief in a religion helps one understand the meaning of life.
5. All things in the universe have been determined.

References

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*, 569–576.

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making, 2*, 81–94.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193*, 31–35.

Audley, R. J. (1960). A stochastic model for individual choice behavior. *Psychological Review, 67*, 1–15.

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning & Memory, 5*, 241–252.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 929–945.

Bassili, J. N. (2003). The minority slowness effect: Subtle inhibitions in the expression of views not shared by others. *Journal of Personality and Social Psychology, 84*, 261–276.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Mahwah, NJ: Erlbaum.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68.

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. *Advances in Experimental Social Psychology, 42*, 319–373.

Bolte, A., & Goschke, T. (2005). On the speed of intuition: Intuitive judgments of semantic coherence under different response deadlines. *Memory & Cognition, 33*, 1248–1255.

BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.

Bortolloti, L. (2008). *An introduction to the philosophy of science*. Cambridge, MA: Polity Press.

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Oxford University Press.

Bovens, L., & Olsson, E. J. (2000). Coherentism, reliability and Bayesian networks. *Mind, 9*, 685–719.

Brewer, W. F., & Sampaio, C. (2012). The metamemory approach to confidence: A test using semantic memory. *Journal of Memory and Language, 67*, 59–77.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*, 26–48.

Carruthers, P. (1992). *Human knowledge and human nature: A new introduction to an ancient debate*. Oxford: Oxford University Press.

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory & Cognition, 22*, 273–280.

Chisholm, R. (1989). *Theory of knowledge* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Clarkson, J. J., Tormala, Z. L., DeSensi, V. L., & Wheeler, S. C. (2009). Does attitude certainty beget self-certainty? *Journal of Experimental Social Psychology, 45*, 436–439.

Conee, E., & Feldman, R. (2001). Internalism defended. *American Philosophical Quarterly, 38*, 1–18.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt, Rinehart, & Winston.

Davidson, D. (1986). A coherence theory of truth and knowledge. In E. LePore (Ed.), *Truth and interpretation: Perspectives on the philosophy of Donald Davidson* (pp. 307–319). Oxford: Basil Blackwell.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality & Social Psychology Review, 14*, 238–257.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*, 274–290.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage.

Engel, P. (1998). Believing, holding true, and accepting. *Philosophical Explorations, 2*, 140–151.

Fazio, R. H. (1979). Motives for social comparison: The construction–validation distinction. *Journal of Personality and Social Psychology, 37*, 1683–1698.

Feldman, R. (2003). *Epistemology*. Upper Saddle River, NJ: Prentice-Hall.

Feltz, A. (2009). Experimental philosophy. *Analyse & Kritik, 31*, 201–219.

Feltz, A., & Cokely, E. T. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness and Cognition, 18*, 342–350.

Feltz, A., & Zarpentine, C. (2010). Do you know more when it matters less? *Philosophical Psychology, 23*, 683–706.

Festinger, L. (1950). Informal social communication. *Psychological Review, 57*, 271–282.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140.

Galton, F. (1907). Vox populi. *Nature, 75*, 450–451.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229–244). London: Oxford University Press.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*, 208–214.

Gendler, T. S. (2008). Alief and belief. *Journal of Philosophy, 105*, 634–663.

Gettier, E. (1963). Is knowledge justified true belief? *Analysis, 23*, 121–123.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance. Knowing that you don't know. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 311–325.

Goldman, A. I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy, 73*, 771–791.

Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge: Harvard University Press.

Goldman, A. I. (1979). What is justified belief? In G. S. Pappas (Ed.), *Justification and knowledge: New studies in epistemology* (pp. 1–23). Dordrecht, Holland: D. Reidel.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.

Hampton, J. A. (2011). Concepts and natural language. In R. Belohlavek & G. J. Klir (Eds.), *Concepts and fuzzy logic*. Cambridge: MIT Press.

Hansen, J., & Wänke, M. (2010). Truth from language and truth from fit: The impact of linguistic concreteness and level of construal on subjective truth. *Personality & Social Psychology Bulletin, 36*, 1576–1588.
Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16*, 107–112.
Huemer, M. (1997). Probability and coherence justification. *Southern Journal of Philosophy, 35*, 463–472.
Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 1–47). New York, NY: Academic Press.
Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review, 104*, 344–366.
Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*, 1–24.
Kirkham, R. L. (1992). *Theories of truth*. Cambridge: MIT.
Klein, P., & Warfield, T. A. (1994). What price coherence? *Analysis, 54*, 129–132.
Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16*, 309–324.
Knobe, J. (2007). Experimental philosophy. *Philosophy Compass, 2*, 81–92.
Kolers, P. A., & Palef, S. R. (1976). Knowing not. *Memory & Cognition, 4*, 553–558.
Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.
Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.
Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9*, 149–171.
Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*, 945–959.
Koriat, A. (2011). Subjective confidence in perceptual judgments: A test of the self-consistency model. *Journal of Experimental Psychology: General, 140*, 117–139.
Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review, 119*, 80–113.
Koriat, A. (in press-a). Confidence in personal preferences. *Journal of Behavioral Decision Making.*
Koriat, A. (in press-b). The subjective confidence in one's knowledge and judgments: some metatheoretical considerations. In M. Beran, J. L. Brandl, J. Perner, & J. Proust (Eds.), *The foundations of metacognition* (pp. 213–233). Oxford: Oxford University Press.
Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science, 13*, 407–551.
Koriat, A., & Adiv, S. (2011). The construction of attitudinal judgments: Evidence from attitude certainty and response latency. *Social Cognition, 29*, 577–611.
Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.
Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483–502). New York, NY: Guilford Press.
Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 107–118.
Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*, 36–69.
Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 117–135). New York, NY: Psychology Press.
Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). New York, NY: Cambridge University Press.
Kurman, J., & Ronen-Eilon, C. (2004). Lack of knowledge of a culture's social axioms and adaptation difficulties among immigrants. *Journal of Cross Cultural Psychology, 35*, 192–208.
Lemos, N. (2007). *Introduction to theory of knowledge*. Cambridge: Cambridge University Press.
Leung, K., Bond, M. H., De Carrasquel, S. R., Muñoz, C., Hernández, M., Murakami, F., et al (2002). *Journal of Cross-Cultural Psychology, 33*, 286–302.
Lewis, C. I. (1946). *An analysis of knowledge and valuation*. La Salle: Open Court.
Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York, NY: Cambridge University Press.
Loussouarn, A., Gabriel, D., & Proust, J. (2011). Exploring the informational sources of metaperception: The case of Change Blindness Blindness. *Consciousness and Cognition*, 1489–1501.
Luus, C. A. E., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79*, 714–723.
Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition, 92*, 1–12.
Markie, P. (2008). Rationalism vs. empiricism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. The Center for the Study of Language and Information (CSLI), Stanford University.
McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory & Cognition, 6*, 462–472.
McKenzie, C. R. M. (1997). Underweighting alternatives and overconfidence. *Organizational Behavior and Human Decision Processes, 71*, 141–160.
Michaelian, K. (2012). Metacognition and endorsement. *Mind & Language, 27*, 284–307.
Moser, P. K. (1988). Internalism and coherentism: A dilemma. *Analyses, 48*, 161–163.
Nichols, S. (2004). After objectivity: An empirical study of moral judgment. *Philosophical Psychology, 17*, 5–28.
Olsson, E. (2001). Why coherence is not truth-conducive. *Analysis, 61*, 236–241.
Olsson, E. (2002). What is the problem of coherence and truth? *Journal of Philosophy, 99*, 246–272.
Olsson, E. (2005). *Against coherence: Truth, probability, and justification*. New York, NY: Oxford University Press.
Orive, R. (1988). Group consensus, action immediacy, and opinion confidence. *Personality and Social Psychology Bulletin, 14*, 573–577.
Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*, 972–987.
Pelham, B. W. (1991). On confidence and consequence: The certainty and importance of self-knowledge. *Journal of Personality and Social Psychology, 60*, 518–530.
Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy (Corrected Ed., 1962)*. Chicago: University of Chicago Press.
Proust, J. (2007). Metacognition and metarepresentation: Is a self-directed theory of mind a precondition for metacognition? *Synthese, 159*, 271–295.
Proust, J. (2008). Epistemic agency and metacognition: An externalist view. *Proceedings of the Aristotelian Society, 108*, 241–268.
Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology, 1*, 563–581.
Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology, 82*, 416–425.
Schervish, M. J. (1996). *P* values: What they are and what they are not. *The American Statistician, 50*, 203–206.
Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feeling of knowing. *Psychonomic Bulletin & Review, 1*, 357–375.
Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14*, 332–348.
Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition, 25*, 638–656.

Schwarz, N., & Strack, F. (1985). Cognitive and affective processes in judgments of well-being: A preliminary model. In H. Brandstatter & E. Kirchler (Eds.), *Economic psychology* (pp. 439–447). Linz, Austria: Taubler.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.
Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology, 53*, 1–26.
Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Anchor Books.
Swain, S., Alexander, J., & Weinberg, J. (2008). The instability of philosophical intuitions: Running hot and cold on true temp. *Philosophy and Phenomenological Research, 76*, 138–155.
Thagard, P. (2002). *Coherence in thought and action*. Cambridge: MIT Press.
Thagard, P. (2009). Why cognitive science needs philosophy and vice versa. *Topics in Cognitive Science, 1*, 237–254.
Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition, 18*, 22–38.
Van Cleve, J. (2005). Why coherence is not enough: A defense of moderate foundationalism. In M. Steup & E. Sosa (Eds.), *Contemporary debates in epistemology* (pp. 168–180). Oxford: Blackwell.
Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 582–600.
Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics, 13*, 37–58.
Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*, 1049–1062.
Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 37–65). Hillsdale, NJ: Erlbaum.
Wright, J. C. (2010). On intuitional stability: The clear, the strong, and the paradigmatic. *Cognition, 115*, 491–503.
Zakay, D., & Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychologica, 98*, 103–115.