Supporting Text

1. Comparison of Performance in the Replication and Questionnaire Conditions.

It is important to assess whether the questionnaire had a significant impact on participants' behavior. Fig. 4 shows that the number of participants who selected from one of the two best decks on each trial is quite similar for the replication and questionnaire conditions. (As mentioned in the paper, the best decks are determined based on each participant's experience with the decks up until the trial in question.) An interesting difference is that in the questionnaire condition there is typically a peak in the trial right after the questions (i.e., trials 21, 31, and so on), whereas that does not occur in the replication condition. We believe that such peaks reflect an influence of the questionnaire on the selection in the immediately following trial. This may be due to a tendency for participants to select the deck that they just identified as the best deck in the last question in the questionnaire. Interestingly, however, this influence does not seem to carry over to the other trials, given that the general profile of selections is comparable across conditions. In short, although it seems that the questionnaire may have some influence on participants' behavior, particularly in the trial immediately following question periods, such influence seems to be quite small overall.

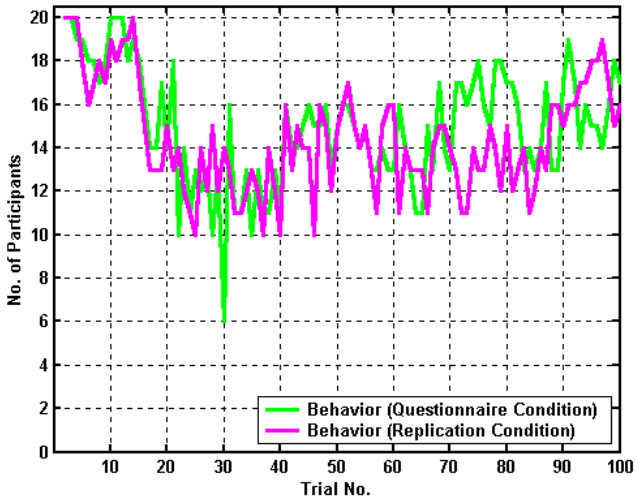

Fig. 4. Number of participants who selected from one of the best two decks on each trial, for the questionnaire and replication conditions.

One arrives at the same conclusion if comparing the number of selections from "good" and "bad" decks across conditions, using Bechara *et al.*'s (1) definition of good and bad decks. (Recall that, according to Bechara *et al.* (1), the good decks are the decks that give only $50 but lead to net gains and the bad decks are the decks that give $100 but lead to net losses.) Specifically, the difference in total number of cards selected from the $50 and $100 decks for the condition that included our questionnaire and the condition that replicated the study by Bechara *et al.* was not

statistically significant. The mean number of total selections from the good decks was 63.55 at the end of the game for the questionnaire condition and 58.6 for the replication condition. The difference of 4.95 cards was not significant (confidence interval: -4.11 to 14.01; two-sided *P* value = 0.28). This difference was also not significant at any point in the game. Thus, again, if there was an impact of the questionnaire on participants' performance, it seems to have been slight.

In summary, the questionnaire seems to have had little impact on participants' behavior. We take this to be suggestive evidence that the questionnaire did not appreciably change the way participants perceived and approached the game.

2. Verbal Reports of Participants in the Replication Condition.

In this section, we describe our analysis of the verbal reports of participants in the replication condition. Our goal was to determine whether, using the same procedures as Bechara *et al.* (1), we would also find in our population that participants would behave advantageously before knowing the advantageous strategy.

In their study, Bechara *et al.* (1) classified participants' explicit knowledge of the game into three categories: "pre-hunch," "hunch," and "conceptual period." (There was a fourth category, "pre-punishment," which corresponded to the period before participants received any punishment and always occurred before the pre-hunch period. In most of our analyses, it is subsumed in the pre-hunch period.) In the main text of their article, Bechara *et al.* (1) discuss these periods as follows:

> By about card 50, all normal participants began to express a "hunch" that decks A and B [which are the "bad" decks] were riskier […]. We called this period hunch. […] By card 80, many normal participants expressed knowledge about why, in the long run, decks A and B were bad and decks C and D were good. We called this period conceptual. (Ref. 1, p. 1293.)

Further detail is provided in a figure caption on the next page:

> The pre-hunch period consisted of the […] series of cards when subjects […] professed no notion of what was happening in the game [...]. The hunch period […] corresponded to the period when subjects reported "liking" or "disliking" certain decks, and "guessed" which decks were risky or safe, but were not sure of their answers […]. The conceptual period corresponded to the period when subjects were able to articulate accurately the nature of the task and tell for certain which were the good and bad decks, and why they were good or bad. (Ref. 1, p. 1294.)

These definitions raise several difficulties. First, it is not clear how to classify participants that profess some knowledge about the game, if such knowledge is incorrect. It would seem that such participants could not be classified as being in the pre-hunch period, as they do "profess [some] notion of what was happening in the game." However, it is also not clear that they could be in the hunch period, given that they do not "express a 'hunch' that decks A and B were riskier." Second,

it is not clear how to classify participants that express a correct hunch or knowledge, if these do not specify that both decks C and D are good (or, equivalently, that both decks A and B are bad). In this category, it is necessary to distinguish two cases. In one case, participants have a correct hunch or knowledge about the game, but such hunch or knowledge would not be sufficient to guide advantageous behavior. An example is a correct hunch or knowledge that one of the bad decks is bad. Such knowledge would be sufficient to increase the probability of advantageous behavior by eliminating one of the decks from consideration, but it would not be sufficient to ensure advantageous behavior, because the other bad deck could still be selected. In the other case, a participant expresses a hunch or knowledge that one of the good decks is good. Such knowledge is sufficient to guide advantageous behavior, because the participant could play exclusively from that deck. In our data, this situation arises frequently, as it is common for participants to hone in on one of the good decks and not explore the other very much.

To address these issues, we refined the categories of Bechara *et al.* (1). Overall, we had six categories: no professed knowledge, incorrect or incomplete hunch/knowledge, partial hunch, hunch, partial conceptual, and conceptual. Participants who professed no notion of what was going on in the game were assigned to the "no professed knowledge" category. Participants who did profess some hunch or knowledge about the game, but for whom such hunch or knowledge was not sufficient to guide advantageous behavior, were assigned to the "incorrect or incomplete hunch/knowledge" category. Participants in this category might have expressed an incorrect hunch or knowledge, a hunch or knowledge that one of the bad decks was bad in the absence of further information, or other type of correct knowledge about the game that was nevertheless insufficient to guide advantageous behavior. Participants who had a hunch that one of the good decks was good were assigned to the "partial hunch" category, and those that had a hunch that both good decks were good were assigned to the "hunch" category. Participants who had conceptual knowledge that one of the good decks was good were assigned to the "partial conceptual" category, and those that had conceptual knowledge that both good decks were good were assigned to the "conceptual" category.

Note that knowing which two decks are good is equivalent to knowing which two decks are bad. In our data, it is sometimes the case that participants refer only to the good (or bad) decks when stating their knowledge about the game, but our impression was that cases like this generally reflect an understanding of this complementarity. We therefore did not require complete mention of all decks; in other words, having a hunch or knowledge that both bad decks were bad was treated equivalently to having a hunch or knowledge that both good decks were good. Note that having a hunch or knowledge that just one of the bad decks is bad is not complementary to having a hunch or knowledge that one of the good decks is good. The former is not sufficient to guide advantageous behavior and would be classified in the "incorrect or incomplete hunch/knowledge" category. The latter, on the other hand, is sufficient to guide advantageous behavior and would be classified in the "partial hunch" or "partial conceptual" category (depending on whether the participant had verbalized a hunch or conceptual knowledge about the good deck being good).

To ensure that it was possible to unequivocally classify participants' verbal reports into these categories, we devised a decision tree for the categorization of the knowledge participants have about the game (Fig. 5). To simplify the graphical display of our decision tree, we focused the descriptions in the tree on the good decks, with the understanding that a correct hunch or knowledge about both bad decks being bad should be treated as equivalent to a correct hunch or knowledge about both good decks being good.

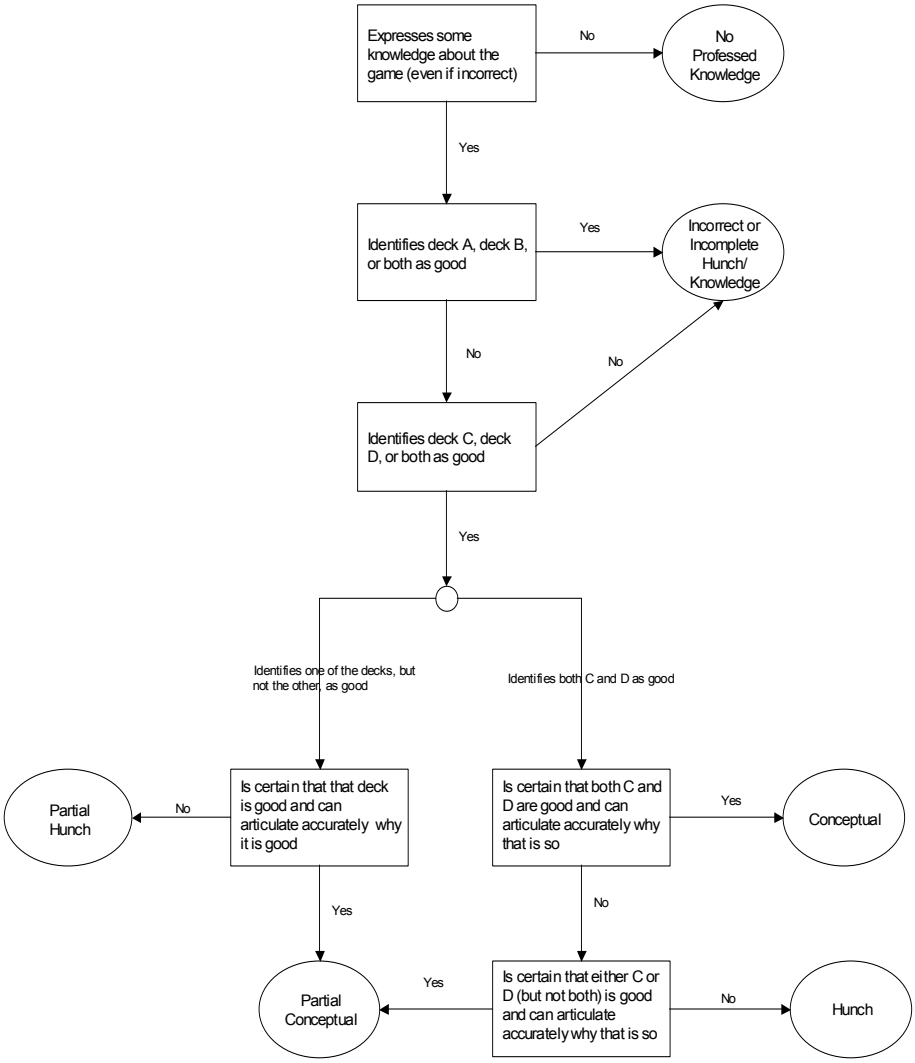

Fig. 5. Decision tree for the categorization of participants' knowledge.

Three naïve raters, blind to our hypotheses and the purpose of our experiment, classified the knowledge of participants in the replication condition using the decision tree in Fig. 5. The three

raters were upper-level undergraduates at Carnegie Mellon University, without psychology or social and decision sciences background. The raters were provided with the decision tree in Fig. 5 and instructed to treat a correct hunch or knowledge about both bad decks being bad as equivalent to a correct hunch or knowledge about both good decks being good. During a training session, raters had to classify a small set of made-up examples to ensure that they were able to use the decision tree accurately. These examples included training in all branches of the decision tree, and also illustrated the fact that knowledge about both decks A and B being bad should be considered equivalent to knowledge about both decks C and D being good. Raters had to get all training examples right before they were given the actual participant data.

To be able to compare our results with those of Bechara *et al.* (1), a policy is required to group our categories into those of Bechara *et al.* (1). Table 1 shows the correspondences we used. As noted above, Bechara *et al.* define the conceptual period as "the period when subjects were able to articulate accurately the nature of the task and tell for certain which were the good and bad decks, and why they were good or bad" (1). This suggests that Bechara *et al.* (1) would treat situations that fell in our partial conceptual category as not belonging to the conceptual period – and hence, as belonging to the hunch period. Nevertheless, to be as flexible as possible in our interpretation of the way Bechara *et al.* (1) categorized their participants, we grouped our partial conceptual category with the other categories in two different ways. Under one grouping, which we call "partial," participants who were in our partial conceptual period (i.e., those who knew that one of C or D, but not both, was good and why it was good) were assigned to Bechara *et al.*'s (1) conceptual period. Under the other grouping, which we call "both," participants had to be in our conceptual period (i.e., they had to know that both C and D were good and why they were good) to be classified as conceptual according to the categories of Bechara *et al.* (1); our partial conceptual category was therefore assigned to Bechara *et al.*'s hunch category. We will focus our analyses on the "both" grouping, because it seems to reflect the criterion that Bechara *et al.* (1) laid out for classifying participants as being in the conceptual period. However, we will also show that our conclusions would not be changed under the other grouping.

In their paper, Bechara *et al.* (1) also suggest that knowledge is progressive and that participants always progress from the pre-hunch period to the hunch period and then to the conceptual period (although some participants never reach the conceptual period). However, all three of our raters found that often participants seem to have the right hunch and then wrongly revise their hypothesis, or seem to have understood everything conceptually but then go back to believing something wrong or to being uncertain about their knowledge. Bechara *et al.* (1) do not discuss how such cases should be handled. To be able to compare our results with theirs, we again took two contrasting approaches, intended to span the range of possible approaches that Bechara *et al.* (1) may have used. In one approach, which we call conservative, participants were considered to be in a particular state of knowledge (i.e., to be in the hunch or conceptual period) only if they never went back to a lower state of knowledge. In the other approach, which we call aggressive, participants were considered to have entered a particular period the first time that they were rated as being in that period. Consider, for example, the case of a participant who was classified as

being in the following states of knowledge in each question period, after we transformed our categories into those of Bechara *et al.* (1): pre-hunch, pre-hunch, hunch, pre-hunch, hunch, hunch, conceptual, hunch, and conceptual. In the conservative approach, this participant would be considered to have entered the hunch period only at the fifth assessment (trial 60) and to have entered the conceptual period only at the last assessment (trial 100). In the aggressive approach, this participant would be considered to have entered the hunch period at the third assessment (trial 40) and the conceptual period at the seventh assessment (trial 80).

Our category	Bechara *et al.*'s category
No professed knowledge	Pre-hunch
Incorrect or incomplete hunch/knowledge	Pre-hunch
Partial hunch	Hunch
Hunch	Hunch
Partial conceptual	In the "both" grouping: Hunch In the "partial" grouping: Conceptual
Conceptual	Conceptual

Table 1. Correspondence between our categories of knowledge and those of Bechara *et al.* (1).

Table 2 shows the results of Bechara *et al.* (1). As can be seen, in their population of normal participants, all reached the hunch period, but 30% did not reach the conceptual period. Participants entered the hunch period on average on trial 50, and of those who reached the conceptual period, they did so on average on trial 80. Table 3 shows the results of our analyses for the "both" and "partial" groupings, under the conservative and aggressive approaches. The results presented correspond to the average over the three raters, with the standard deviation shown in parentheses.

The results in Table 3 show that under the "both" grouping, the aggressive approach has the closest resemblance to the results of Bechara *et al.* (1). Under this analysis, 12% of our participants did not reach the hunch period, compared with 0% in the Bechara *et al.* study (1), and of those who did, the average trial number in which they did so was 43, compared with 50 in the Bechara *et al.* study. We had significantly more participants that never reached the conceptual period than Bechara *et al.* (1) (60%, compared with 30% in their study), but of those who did, they did so on average slightly earlier (on trial 72, compared with trial 80 in the Bechara *et al.* study).

	Bechara *et al.*'s results
Percentage of participants who did not reach the hunch period	0
For participants who reached the hunch period, average trial number in which they did so	50
Percentage of participants who did not reach the conceptual period	30
For participants who reached the conceptual period, average trial number in which they did so	80

Table 2. Results from Bechara *et al.* (1).

		Conservative approach	Aggressive approach
	Percentage of participants who did not reach the hunch period	37 (5.8)	12 (2.9)
	For participants who reached the hunch period, average trial number in which they did so	62 (5.8)	43 (4.6)
Partial grouping	Percentage of participants who did not reach the conceptual period	47 (7.6)	25 (8.7)
	For participants who reached the conceptual period, average trial number in which they did so	74 (1.6)	62 (6.6)
Both grouping	Percentage of participants who did not reach the conceptual period	77 (5.8)	60 (5)
	For participants who reached the conceptual period, average trial number in which they did so	91 (5.6)	72 (4.8)

Table 3. Results from our analyses, averaged over the three raters, with standard deviations in parentheses.

Fig. 6 shows the mean number of card selections from good and bad decks in the several periods (pre-punishment, pre-hunch, hunch, and conceptual), under this analysis. (This figure is homologous to the bottom left panel of figure 1 in ref. 1.) As can be seen in the figure, in the hunch and conceptual periods participants select more from the good decks than from the bad decks. Such differences are statistically significant (in both periods, $P < 0.01$ in paired t tests). Unlike Bechara *et al.* (1), we did not find that participants selected more from the good decks during the pre-hunch period. However, the difference that they found was small and it was not statistically significant.

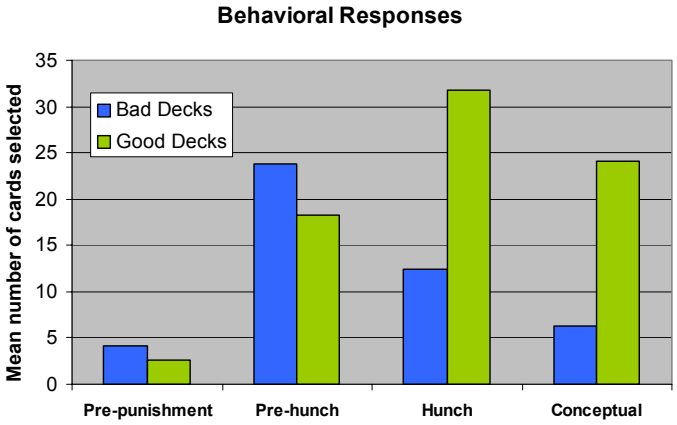

Fig. 6. Mean number of card selections from good and bad decks during the different knowledge periods, for the "both" grouping/ aggressive approach.

In summary, the application of Bechara *et al.*'s (1) data analysis strategy to our participant population replicated their statistically significant results of advantageous behavior in the hunch and conceptual periods. This means that it is unlikely that our findings in the questionnaire condition are due to our participants in general having more explicit knowledge than theirs. In fact, since fewer participants in our population reach the conceptual period, if anything, it would seem that our participants have less explicit knowledge than those in their study.

While the "both" grouping with the aggressive approach is the analysis that most closely fits what Bechara *et al.* (1) report that they have done, even under the most liberal overall analysis – the partial grouping with the aggressive approach – our conclusions would not change. Indeed, even under that analysis, participants selected more from the good decks than the bad decks in both the hunch ($P < 0.01$) and the conceptual ($P < 0.001$) periods. Note that if, as seems likely, Bechara *et al.* (1) actually used stricter standards than those in this analysis, the conclusion that our participants also behaved advantageously before knowing the advantageous strategy would be even more strongly supported – as the data in Table 3 indicates, with the enforcement of stricter standards, our participants enter the hunch and conceptual periods later and less frequently.

3. The Behavior of Participant 41.

As mentioned in the main text, participant 41 sometimes behaved advantageously but did not show knowledge of the advantageous strategy in one or both of questions Q1 and Q5. However, this participant seems to have behaved randomly during a substantial portion of the game. Fig. 7 shows his behavioral selections throughout the game. As can be seen, starting on trial 31, his selections seem to vary randomly across all decks, suggesting that he did not develop a consistent preference for the good decks. Indeed, from trials 31 to 100, this participant selected deck A 17 times, deck B 18 times, deck C 16 times, and deck D 19 times, which is not significantly different from chance behavior ($\chi^2 = 0.2857$ with 3 d.f.; $P = 0.96$). It is noteworthy that on trial 30 this participant got a second $1,250 loss on deck B, from which he was consistently playing. It seems that at that point he gave up on trying to understand the game and decided to play randomly.

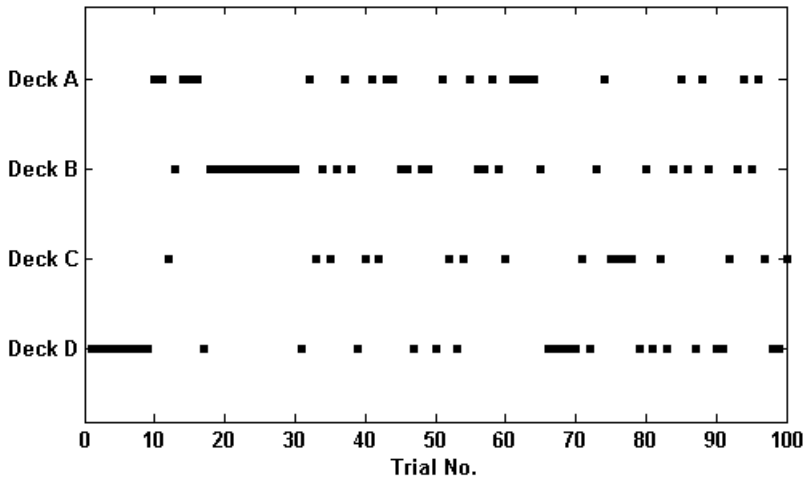

Fig. 7. The behavioral selections of participant 41 throughout the game.

Given this random behavior starting on trial 31, it is not surprising that on some trials this participant behaved advantageously but failed to show evidence of knowledge of the advantageous strategy in one or more of the measures of Level 1 or Level 2 knowledge; in all likelihood, this occurred on trials in which the participant behaved advantageously by chance. Indeed, all cases in which this participant behaved advantageously and failed to show knowledge of the advantageous strategy in at least one of the Level 1 or Level 2 measures occurred on or after trial 40, by which time the participant already seems to be behaving randomly.

4. The Behavior and Verbal Reports of Participant 36.

As mentioned in the main text, on trials 40, 70, and 80, participant 36 behaved advantageously but did not show knowledge of the advantageous strategy in either of the quantitative measures of Level 2 knowledge – the reported net or the calculated net. Here, we analyze this participant's answers to question Q2, in order to determine whether those provide evidence that this participant

had *qualitative* Level 2 knowledge that could have provided a basis for her advantageous behavior on those trials.

It is important to note from the outset that Q2 suffers from the same problems and limitations as the original questions by Bechara *et al.* (1). The answers to Q2 are therefore likely to be at best pale reflections of the qualitative Level 2 knowledge that participants actually have. In fact, it is noticeable that participants' answers to Q2 tend to become shorter in successive question periods, and this was certainly the case for participant 36 (Fig. 8). Although it may be argued that participants may have more knowledge to convey in the earlier question periods, our sense from running the experiment is that when participants realize that these questions may come up several times, they seem eager to give short, quick, and often easy answers, presumably to avoid prolonging the experiment too much. (Recall that these are undergraduate students who are participating in the experiment for credit.)

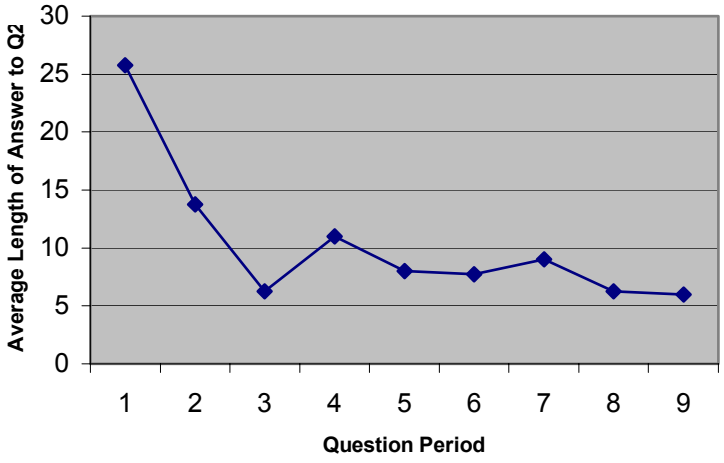

Fig. 8. Average length (in number of words) of the answers of participant 36 to question Q2, by question period. Note that by the third question period, answers are much shorter than in the beginning.

There are also specific problems with question Q2. The answer to the question of why one gives a particular rating to a deck necessarily relates to the observed values for the other decks (e.g., a deck that has an expected net result of $25 seems much worse if two other decks have expected nets of $100 than if those two decks have expected nets of $–25); it further relates to the ratings that one gives to the other decks. Moreover, a participant may answer question Q2 from many different perspectives – for example, focusing on why he or she did not give the deck a higher rating, on why he or she did not give the deck a lower rating, on why he or she gave the deck a rating that is larger/smaller than the rating he or she gave to another deck, and so on. More generally, Q2, like other open-ended questions, is under-constrained, and as such is very unlikely to give a full picture of participants' knowledge.

Despite all these problems with question Q2, it is informative to look at the answers of participant 36 to this question, because they do show evidence of qualitative knowledge that would seem sufficient to guide her ratings and advantageous behavior on those trials in which her answers to the quantitative Level 2 questions do not reflect such knowledge (40, 70, and 80).

Table 4 shows details of the behavior and verbal reports of participant 36 throughout the game. Note that in our study, unlike in the original study of Bechara *et al.* (1), the positions of the decks were randomized and the decks were not explicitly labeled with a letter. Participant 36 had the decks in the order A, D, C, B, and refers to them by their position number (e.g., 'deck 2' would mean deck D).

From participant 36's answers to Q2, it seems that by trial 40 she has understood that deck 1 (deck A) is a bad deck – on trials 30 and 40, she rates this deck with a –2 "because I lost twice in a short period of time" (trial 30) and that deck "still provides a large loss" (trial 40), and indeed she will only play from this deck twice more until the end of the game. She also seems disappointed with deck 2 (deck D), misremembering the fact that she had a large loss ($250) as having lost often. She seems quite happy with deck 3 (deck C), giving it a very large rating (8), because she "[hasn't] really lost that much" in that deck. (Indeed, up until trial 40, she never has a net loss on deck C.) She still seems to believe that deck 4 (deck B) may be good "because the winnings are still pretty high," even though she seems aware of having had a large loss there – on trial 20, when talking about her loss of $150 on deck 1, she said "the losing has only been $150, compared to deck 4, which was a lot worse." Given this state of knowledge, it hardly seems surprising that she selects mostly from deck C both in the period leading up to the question period on trial 40 and in the following period (and indeed for the rest of the game). Furthermore, given that on trial 40 she still seemed to think that deck B was a good deck ("because the winnings are still pretty high"), it also does not come as a surprise that she went back to deck B in the next period (only to lose $1,250 again). In fact, this pattern is not at all unusual. It often takes two $1,250 losses in deck B to convince participants to stay away from that deck. Even then, many (including participant 36) still go back to that deck a couple of times – sometimes because they understand that there is usually a series of "safe" $100 wins before the next $1,250 loss, and other times because they are higher risk-takers.

Between trials 40 and 70, she plays 26 out of 30 times from deck C, which is in line with her stated knowledge about that deck: "when I lose, I don't lose that much, and sometimes the wins and the losses cancel out" and "the losses aren't huge and I usually win or… you know…" (which we interpret as "I usually win or stay even"). (In fact, most participants find it easy to figure out that deck C provides them with a net gain. This is not surprising, given the structure of that deck. Typical card sequences in that deck are a gain of $50, then a gain of $50 and a loss of $50, then a gain of $50 again, then a gain of $50 and a loss of $50, and so on. It is therefore often easy for participants to see that they typically win or stay even on that deck.) At the same time, she seems to have come to understand that deck D is also a good deck, "because I win a lot of times and don't lose very often." Indeed, after trial 70 and until the end of the game, she plays

mostly from decks C and D (16 times and 11 times, respectively: a total of 27 times out of 30). Starting on the question period on trial 50 – that is, after having gotten the second $1,250 loss – it is also clear to her that deck B is a bad deck ("because when I lose, a lose a lot"). It is interesting to note that starting on trial 50, she seems to forget that deck A was a bad deck. Indeed, by trial 70, she indicates considerable uncertainty about this deck, saying "I don't use it very often, so I don't really know," and by trial 80 she is claiming that in that deck "the number of wins and losses balances out." Equally interesting is that, consistent with this misunderstanding, she does seem to go back to exploring deck A in the last period of the game, but is soon discouraged by a $200 loss, giving that deck a negative rating in the last question period ("because the losses are large").

It is important to note that, in the case of this participant, the quantitative measures of Level 2 knowledge do not reflect all of the qualitative knowledge that emerges from an analysis of the answers to Q2. For example, on trial 40, she rated deck A with a –2, "because it still provides a large loss," while at the same time her reported net and calculated net for that deck were positive (50 and 60, respectively). A similar pattern occurs in her responses to the questions about deck A on trial 100. We note in this regard that, whereas for the great majority of participants the notion of an average did not pose any challenge, there were at least one or two participants to whom we had to explain the concept of an average, and it was not entirely clear that they fully understood the difference between an average and a mode, even after a fairly detailed explanation. Unfortunately, we did not make a special note of these participants at the time and cannot now reproduce whether participant 36 was in that situation. In any case, it is not at all implausible that participant 36 may have had trouble with the concept of an average, which would limit her ability to convert her Level 2 qualitative knowledge into the averages required to answer questions Q3.1 and Q3.4. This could explain why, in her case, the reported net and the calculated net did not show evidence of Level 2 knowledge.

To summarize, the kind of data obtained in answer to question Q2 is, by its own nature, scanty and more open to interpretation. Nevertheless, the analyses above lend support to the idea that, on trials 40, 70, and 80, participant 36 had Level 2 qualitative knowledge that would provide a sufficient basis to guide her ratings and behavior. It therefore appears that, even for participant 36, when she behaved advantageously, she had Level 2 knowledge.

Question period		Deck 1 (deck A)	Deck 2 (deck D)	Deck 3 (deck C)	Deck 4 (deck B)
Trial 20	*Deck outcomes*	100, 100, 100 - 150	50, 50, 50	50, 50, 50 – 50	100, 100, 100, 100, 100, 100, 100, 100, 100 – 1250, 100, 100
	Actual net	50	50	33.33	-13.64
	Rating	8	2	0	5
	Why?	Because the winnings have been $100 and the losing has only been $150, compared to deck 4, which was a lot worse.	Because it wasn't a bad deck, but the wins were only $50 and I haven't even pressed it that much, so I assume it's about a 2.	Because it was just a little bit worse than deck 2. I have lost money on that deck [deck 3], I think, $50.	Because it seems good because I won a lot of money in the beginning and then all of a sudden I lost, but it seemed like you could win a lot of money.
	Reported net	70	50	50	50
	Calculated net	80	50	70	0
Trial 30	*Deck outcomes*	100, 100 – 300	50, 50, 50, 50, 50, 50		100, 100
	Actual net	10	50	33.33	3.85
	Rating	-2	8	6	0
	Why?	Because I lost twice in a short period of time.	Because I only won from that deck, I haven't lost.	Just remembering from last time, I only lost a little bit once.	Because the winnings are $100 and that's greater than all the other piles, and I don't lose as often as with deck 1.
	Reported net	10	50	40	25
	Calculated net	-10	50	40	20
Trial 40	*Deck outcomes*		50-250	50, 50 – 50, 50, 50 – 50, 50, 50 – 50, 50 – 50, 50, 50 - 25	
	Actual net	10	25	27.08	3.85
	Rating	-2	3	8	2
	Why?	Because it still provides a large loss.	Because I have lost often.	I haven't really lost that much.	Because the winnings are still pretty high.
	Reported net	50	20	40	80
	Calculated net	60	10	47.5	0

Question period		Deck 1 (deck A)	Deck 2 (deck D)	Deck 3 (deck C)	Deck 4 (deck B)
Trial 50	*Deck outcomes*		50	50 – 75, 50, 50, 50, 50 – 25, 50 – 75, 50, 50 - 50	100 - 1250
	Actual net	10	27.27	25	-78.57
	Rating	1	5	7	-3
	Why?	Because I haven't lost many times.	Because I win a lot of times and don't lose very often.	Because when I lose, I don't lose that much, and sometimes the wins and the losses cancel out.	Because when I lose, I lose a lot.
	Reported net	40	20	40	-200
	Calculated net	-40	20	30	-500
Trial 60	*Deck outcomes*			50, 50, 50, 50 – 50, 50 – 25, 50 – 50, 50, 50, 50 – 75	100
	Actual net	10	27.27	25.86	-66.67
	Rating	2	3	6	-5
	Why?	I haven't lost very often.	I didn't lose often in that deck either.	The losses aren't huge and I usually win or… you know...	Because I lose a lot when I lose.
	Reported net	20	30	40	-150
	Calculated net	10	20	37.5	-500
Trial 70	*Deck outcomes*		50	50 – 50, 50, 50, 50, 50 – 25, 50 – 25, 50, 50 – 75, 50	
	Actual net	10	29.17	26.97	-66.67
	Rating	2	4	5	0
	Why?	Because I don't use it very often, so I don't really know.	Because I don't lose very often.	Because I don't lose that much.	Because I don't use it very often.
	Reported net	70	20	25	-100
	Calculated net	70	10	30	-300

Question period		Deck 1 (deck A)	Deck 2 (deck D)	Deck 3 (deck C)	Deck 4 (deck B)
Trial 80	*Deck outcomes*		50, 50, 50, 50	50 – 50, 50 – 75, 50, 50, 50 – 50, 50	
	Actual net	10	34.38	26.14	-66.67
	Rating	0	5	6	-2
	Why?	Because the number of wins and losses balances out.	I don't lose very often.	Because even though I lose kind of often, I don't lose that much.	Because when I did lose, I lost a lot.
	Reported net	80	30	30	0
	Calculated net	70	30	35	-80
Trial 90	*Deck outcomes*		50, 50, 50, 50 - 250	50 – 50, 50, 50 – 50, 50, 50 – 50, 50 – 50	
	Actual net	10	25	25	-66.67
	Rating	2	3	4	0
	Why?	Because the winnings are $100.	I don't lose very often.	Because I don't lose a lot of money.	Because I lose a lot of money.
	Reported net	75	10	25	-200
	Calculated net	10	10	30	-140
Trial 100	*Deck outcomes*	100, 100 – 200	50, 50, 50	50, 50 – 25, 50 – 75, 50	100
	Actual net	7.14	28.26	25	-56.25
	Rating	-2	0	2	-4
	Why?	Because the losses are large.	Because the losses and the wins balance out.	Because I don't lose a lot.	Because the losses are large.
	Reported net	50	80	10	-200
	Calculated net	60	40	20	-140

Table 4. The behavior and verbal reports of participant 36 throughout the game. The leftmost column indicates the question period. "*Deck outcomes*" indicates the outcomes for each deck in the period immediately preceding that question period. "*Actual net*" indicates the actual observed mean net result for each deck, taking into account the outcomes from the beginning of the game until right before the questions were asked. "*Rating*" indicates the rating that the participant gave to each deck. "*Why?*" indicates the participant's answer to question Q2 for each deck. "*Reported net*" indicates the participant's answer to question Q3.1. "*Calculated net*" indicates the net calculated from the participant's answers to Q2 – Q4.

5. Additional Comments Regarding Fig. 3.

An issue that arises when relating the number of participants who behave advantageously to the number of participants whose verbal reports show evidence of knowledge of the advantageous strategy is whether one should consider behavior in the trials immediately preceding the questionnaire (i.e., trials 20, 30, etc., which we will refer to as trials 10*X*) or the trials immediately following it (i.e., trials 21, 31, etc., which we will refer to as trials 10*X* + 1). Participants have the same experience with the game when answering the questions and when making their selection in the trial immediately following the questionnaire. This would argue for considering behavior in trials 10*X* + 1. However, as pointed out in section 1 above, there is typically a peak in the number of participants who behave advantageously on these trials. In all likelihood, this reflects a direct influence of the questionnaire on behavior in those trials, not the normal course of behavior. We therefore prefer comparing the verbal reports with behavior in the trial immediately preceding the questionnaire (i.e., trials 10*X*). This is what is shown in Fig. 3.

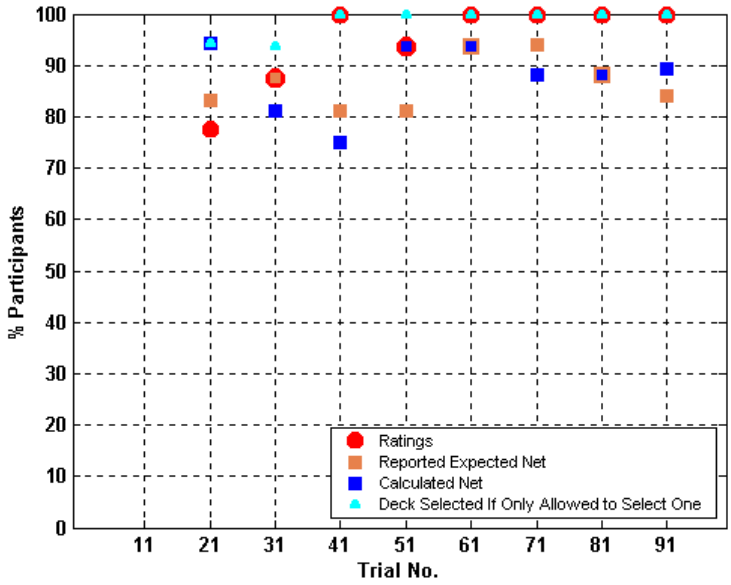

Fig. 9. This figure is similar to Fig. 3, but considering the participants that behaved advantageously on trials 10*X* + 1 instead of trials 10*X*. See text and the legend for Fig. 3 for details.

At the same time, it is also important to consider what would happen if we were to compare the verbal reports with behavior in the trials 10*X* + 1. Recall that we analyze advantageous behavior or verbal reports by considering the outcomes up until the time in which a decision is being made or a report is being produced. This means that when determining if behavior is advantageous on trial 10*X* we consider the outcomes up until, but not including, trial 10*X*, and when determining if a verbal report exhibits knowledge of the advantageous strategy we consider the outcomes up to and including trial 10*X*. Clearly, the outcome of trial 10*X* may have changed a participant's

perception of the game significantly, and one may wonder if our results and conclusions would change if we compared the verbal reports with behavior in trials $10X + 1$ rather than $10X$. In fact, the pattern of results is extremely similar for trials $10X + 1$ (Fig. 9) and $10X$ (Fig. 3). Thus, whereas for simplicity reasons in the paper we focus on one particular way of comparing the verbal reports with behavior, nothing we say hinges on that.

6. Participants' Certainty that They Know What They Should Do in the Game.

Fig. 10 shows the mean of participants' answers to question Q4 per question period. As can be seen, participants' certainty that they know what they should do in the game increases monotonically as the game progresses (seeming to level off in the last two question periods).

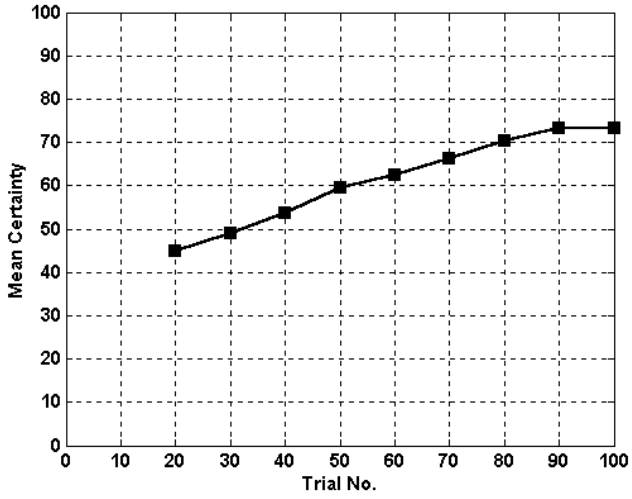

Fig. 10. Mean certainty reported by participants in answer to question Q4 per question period.

1. Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. (1997) *Science* **275,** 1293–1295.