Research Article

THE PRICE OF EXPERTISE: Effects of Experience on the Water-Level Task

Heiko Hecht[1] and Dennis R. Proffitt[2]

[1]NASA Ames Research Center and [2]University of Virginia

Abstract—*When shown a tilted container, people often fail to appreciate that the surface of the liquid contained within should remain horizontal with respect to the ground This study investigated how amenable this bias is to experience in relevant everyday situations Surprisingly, liquid surfaces that waitresses and bartenders considered natural deviated even more from horizontal than was the case for comparison groups This finding is, to our knowledge, the only documented case in which performance declines with experience We suggest that practical experience promotes a functionally relative perspective, in which the orientation of the liquid's surface is evaluated relative to that of its container as opposed to being related directly to the surrounding environment The container-relative perspective, in turn, evokes a perceptual bias that is responsible for the systematic errors observed on this task*

The intuitive physics literature abounds with examples in which people's commonsense ideas conflict with the laws of classical mechanics Regardless of whether such discrepancies are explained by a lack of formal training, the use of mistaken heuristics on the part of the observer (Proffitt & Gilden, 1989), or insufficient attunement to the dynamics of the situation (Runeson & Frykholm, 1983), experience or training is typically assumed to improve performance In this article, we explore a very surprising case in which experience not only fails to improve people's judgments but actually has the opposite effect We found this to be the case with the oft-studied water-level task

THE WATER-LEVEL PROBLEM

The Piagetian water-level task (Piaget & Inhelder, 1948/1956) is one of the most researched problems in cognitive and developmental psychology (Liben, 1991) When asked to indicate the surface orientation of a liquid in a tilted container, young children typically fail to appreciate that the surface remains horizontal with respect to the ground In addition, many adults behave as if they do not know that water remains invariably horizontal regardless of the orientation of its stationary container Up to 40% of the adult population estimates the water level in a container to deviate more than 5° from horizontal (Kalichman, 1988) This failure to appreciate that water remains invariably horizontal is all the more surprising considering that many times a day people observe examples of liquids remaining horizontal, such as whenever they handle a liquid-filled glass

The failure of a substantial proportion of the population to solve the water-level task correctly is quite robust across presentation contexts, and thus does not appear to be an artifact of the technique that is chosen to communicate the task (Howard, 1978, McAfee & Proffitt, 1991) The paper-and-pencil method, which consists of asking subjects to draw the water surface in a depiction of a tilted container, produces errors that are very similar to those found with perceptual versions of the task Howard (1978) presented apparent motion sequences of photographs depicting horizontal and oblique water levels and asked subjects to report whether the sequences represented natural or unnatural events He obtained results similar to those found using paper-and-pencil tasks In another study (McAfee & Proffitt, 1991), performance was similar when observers were presented videotaped displays of a handheld glass containing a dark liquid Tilting the camera and the container during recording produced natural and anomalous water levels These animated versions of the task did not improve performance, even though facilitation is common in a variety of other animations of intuitive physics tasks (Kaiser, Proffitt, Whelan, & Hecht, 1992)

Even though erroneous performance on the water-level task is very robust, there are a large number of modulating influences affecting different subgroups of the adult population Large interindividual differences are consistently found in the water-level task Women are more likely to produce nonhorizontal responses than are men (e g , Kalichman, 1988, Robert & Tanguay, 1990) Biological as well as sociological explanations of the gender effect have been put forth based on findings that high femininity scores obtained in personality tests correlate significantly with the amount of error in the water-level task (Liben, 1991) Also, even though we do not know of any studies that compare college students directly with older people, across studies college-age subjects typically produce fewer errors than do older subjects For example, on average, about 60% of all college students provide judgments of water levels that are within 5° of the horizontal (e g , McAfee & Proffitt, 1991, reported 52 6% correct for female and 79 4% correct for male students) For significantly older subjects, much lower values have been found For instance, subjects between 40 and 72 years of age were tested by Robert and Tanguay (1990) and scored only 31 9% (females) and 39 4% (males) correct All of these subjects were teachers with at least a high school diploma and were thus comparable in their education to college students

One of the most striking aspects of water-level-task performance is the verbal reports by subjects after they finish the task No matter if they are right or wrong, most subjects seem to be quite sure about their answers (Howard, 1978, Liben, 1991) Despite this often unwarranted confidence in their judgments, many subjects reveal a discrepancy between performance and explicit knowledge Roughly one third of subjects

Address correspondence to Heiko Hecht, NASA Ames Research Center, Mail Stop 262-3, Moffett Field, CA 94035-1000, e-mail hecht@eos arc nasa gov

who explicitly know the principle draw nonhorizontal water levels, and one third of those who do not know the principle draw accurate levels (Myer & Hensley, 1984) That is, even though performance and explicit knowledge are positively correlated (Howard, 1978), they are by no means contingent upon each other Thus, erroneous performance on the water-level problem cannot easily be attributed to a bias whose nature is purely conceptual

THE ROLE OF PERCEPTUAL FRAMES OF REFERENCE

An explanation that takes context-specificity into account is that the errors made on the water-level task are due to the manner in which the problem is represented by the subject (McAfee & Proffitt, 1991) As depicted in Figure 1, the problem can be represented in two different ways First, an environment-relative reference system could be adopted, with vertical and horizontal axes derived from the gravitational vertical In this case, if the task were to adjust the water level to horizontal, then the appropriate strategy would be to nullify the discrepancy (α) between the environmental horizontal and the water level Second, an object-relative reference system could be adopted, with the coordinate axes defined relative to the principal axis of the container If the task is represented within an object-relative system, then one must determine what β should be Without knowing precisely what the orientation of the container is relative to the environmental axes, however, the value of β can be determined only qualitatively That is, given that the container has been tilted clockwise, then β indicates how much the water surface must rotate counterclockwise to compensate for the container's tilt

Although the object-relative perspective is less than ideal for solving the water-level problem, it is the representation of choice for most everyday practical situations involving liquids in containers Pick up a full glass of water and notice what spontaneously captures your attention It is the location of the

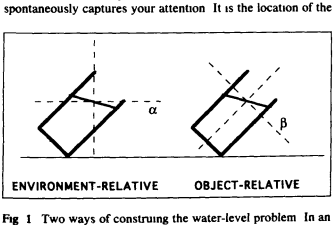

Fig 1 Two ways of construing the water-level problem In an environment-relative reference system (left), the angle α between the water surface and the ground has to be minimized In an object-relative reference system (right), the angle between the water surface and the side of the container has to be close to 90° to avoid spilling the liquid That is, the angle β has to be minimized Thus, a correct horizontal solution to the problem is achieved only when α is minimized

water's surface relative to the rim of the glass This relationship must be controlled if the water is to be prevented from spilling

The adoption of an object-relative perspective, in turn, introduces a perceptual bias that is responsible for the systematic errors made on the water-level problem When environmental orientations are evaluated relative to a local frame of reference that is not in alignment with environmental axes, the environmental axes are misperceived to be slightly reoriented in the direction of the reference frame's tilt The most familiar instance of this bias is the rod-and-frame paradigm (Asch & Witkin, 1948) In this situation, a rod is seen within a tilted frame in an otherwise dark environment The observer is instructed to adjust the rod so as to make it vertical and typically sets the orientation of the rod somewhat off vertical in the direction of the frame's tilt Similarly, when subjects responded to multiple water-level problems, bias in erroneous judgments always reflected settings away from the horizontal in the same direction as the container's tilt (McAfee & Proffitt, 1991) In addition, in a reaction time experiment in which subjects were trained to evaluate whether contrived pictures of liquids in containers had horizontal surface orientations, all subjects required more time to reject pictures in which the surface was inclined off horizontal in the same direction as the container's tilt than to reject pictures in which the surface was inclined opposite to the container's tilt (McAfee & Proffitt, 1991) Thus, the adoption of an object-relative perspective introduces a perceptual bias that is present in everyone The horizontal is misperceived to be oriented away from the horizontal in the direction of the container's tilt

EVERYDAY EXPERIENCE AND THE SELECTION OF FRAMES OF REFERENCE

Although the notion that erroneous performance is a consequence of representing the water-level problem in an object- as opposed to an environment-relative coordinate system is a good description for people's performance, it remains unclear why some people spontaneously pick a frame of reference that produces an adequate solution to the question posed and why others fail to do so We speculate that because practical situations involving liquids in containers promote an object-relative perspective, people whose occupations entail extensive experience with liquid-filled containers might be more inclined to adopt an object-relative perspective On the other hand, the fact that some adults have an accurate conceptualization of the water-level principle whereas others do not could mean that performance on the task is largely unrelated to everyday experience (Howard, 1978) In the present study, we sought to test whether experience does, in fact, play a role in performance, as is predicted by the reference-system notion

There are three possible relationships between experience and performance on the water-level task First, people who have extensive experience involving water levels could exhibit the best performance, as predicted by theories of perceptual learning and attunement (Runeson & Frykholm, 1983) Second, performance on the water-level task and experience with liquid-filled containers may be unrelated because, after all, most adults have considerable experience with such containers

Third, by our account, if the observer's preoccupation not to spill liquids has an effect on the perspective that is taken, then such experience should have a detrimental effect on performance That is, perceptual experience may have an adverse effect on performance in the water-level task This experience consists of controlling the orientation of liquid surfaces with respect to the rims of the containers in a functional context that requires an object-relative perspective To our knowledge, no theory of experience or learning would make such a prediction

OVERVIEW TO STUDIES

To test these three contradicting hypotheses and to find out if performance on the water-level task is related to experience at all, groups having varying levels of relevant experience were tested on a paper-and-pencil version of the water-level task If functionally motivated everyday behavior has an impact on the reference system people use when confronted with the water-level task, their errors should increase with the level of experience they have acquired in such functional contexts Testing subjects who have different levels of experience requires a between-subjects design, which is liable to a number of potential confounds First, age and experience are usually correlated, so we had to select professions that assured different degrees of experience with surface orientation while controlling for age effects Second, age and education are correlated, and young adults tend to perform better than older people on the water-level task Third, education and experience, at least in some professions, are correlated Finally, across studies, gender effects have been found to be very strong in this task for young and old adults (Kalichman, 1988, Robert & Tanguay, 1990) and less strong for middle-aged people Thus, to control for confounds, the following six groups, varying in gender, age, and experience, were tested female and male students, waitresses and housewives, and bartenders and bus drivers To assess performance on the water-level task, paper-and-pencil tests were used in all cases

METHOD

Subjects

Six different groups, each representing a different subpopulation in terms of age, gender, education, and experience, were tested A total of 120 subjects (20 per group) participated in the experiment None of them was familiar with the water-level task

Group SM consisted of male graduate students at Ludwig-Maximilians-Universität München Half of them were enrolled in natural science programs, the other half in the social sciences

Group SF consisted of female students chosen according to the same criteria as their male counterparts The average age of the 40 male and female students was 25 years

The members of Group W were 20 professional waitresses employed at a local brewery servicing the annual Oktoberfest in Munich Their mean age was 43 years Only waitresses with more than 5 years of experience were included in the study Their job at the Oktoberfest consists of picking up full beer mugs and carrying them to thirsty customers, who often are seated more than 50 m away from the keg The mugs hold 1 L of beer and are made of glass so that the surface level of the beer can be seen The waitress has to carry up to five mugs in each hand while maneuvering through the crowded hall Needless to say, delivering the mugs without spilling beer is a task that requires considerable practice

Group H consisted of housewives, chosen as a control for the waitress group They were approximately 54 years old on average, and their education was comparable The typical member of this group, currently or in the past, cared for a household with several children and devoted most of her time to the family Only 25% had part-time jobs, which were in nonacademic fields

Male bartenders working at various bars in Munich made up Group B Five or more years of experience were required to participate in the study The 20 bartenders were on average 33 years old

Finally, Group D consisted of 20 male bus drivers (average age of 49 years) of the Munich public transport system This group was picked because it was equated in education to waitresses and bartenders but was not a profession requiring the control of liquids in containers

Stimuli and Procedure

All subjects were tested individually at their work places and received the same set of instructions They were approached between classes (students) or during short breaks in their jobs and asked if they would participate in a very short study concerning people's judgments of liquids They were told that the whole study consisted of answering three questions All subjects participated voluntarily in the study (2 waitresses and 1 bus driver declined to participate) Subjects were given a questionnaire that consisted of three tasks, the water-level problem and two irrelevant questions concerning liquids The latter were added to prevent subjects from thinking that they were given a trick question and from brooding too long over the water-level question

Page 1 of the questionnaire showed a schematic drawing of the water-level task as depicted in Figure 2 The beaker was tilted 50° clockwise from vertical Subjects were verbally instructed to think of the drawing as a glass held perfectly still by an invisible hand so that the water in it would be at rest Then they were asked to draw a line representing the surface of the water given that the surface would touch the point marked at the right side of the glass It was also pointed out that the glass was being held above the table, which was also visible in the drawing, and that the drawing was a side view of the container, so that a single line would be appropriate to indicate the water level

The identical instructions were also printed on top of the page so that subjects could read them over while solving the problem They were given as much time as they needed to draw the surface of the water Then they were asked to turn to the next page, which depicted two glasses of different widths, one filled with water Subjects were asked to draw the water level in

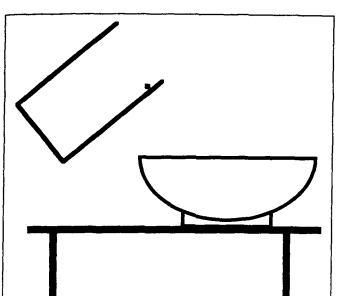

Fig 2 Drawing of the water-level problem presented to all subjects The surface of the water was to be drawn such that it reached the point on the right side of the beaker

the second glass after the contents of the first glass had been poured into it Finally, they had to indicate where the water level in a conelike container would be if it were filled to two thirds of its volume After completing the task, subjects were asked a few questions concerning their educational background, age, and, for the waitresses, years of experience on their jobs

RESULTS

The typical age and gender effects were replicated, and, more important, a clear impact of experience emerged Subjects who were expert handlers of contained liquids showed substantially worse performance on the water-level task

Table 1 shows the frequencies of answers according to three categories Answers were categorized as follows If the surface of the water as drawn by the subject was 5° or less off the horizontal, which is indicated by the dotted line in Figure 3, the answer was rated as correct If the deviation from the horizontal was above 5° in the same direction as the beaker's tilt, a positive bias was marked The solid line in Figure 3 depicts a typical answer in this category ($\alpha = 13°$) Less frequently, subjects drew deviations in the opposite direction (negative deviation)

As shown in Table 1, 65% of the female and 80% of the male students interviewed gave correct answers These numbers correspond roughly to data typically obtained from American college students Thus, the water-level phenomenon does not seem to be culture-specific The nonstudent groups taken together gave 27 5% (females) and 57 5% (males) correct answers The gender effect persists for this population Subjects with more water-level experience (bartenders and waitresses) gave fewer correct answers (32 5%) than the less experienced bus drivers and housewives (52 5%)

The absolute error for each group can be seen in Figure 4 The degrees off horizontal for all subjects in a group were averaged regardless of sign A one-way analysis of variance performed on this measure yielded the following contrasts Males performed significantly better than females, $F(1, 114) = 10\ 66$, $p < 002$ Waitresses and bartenders taken together made larger errors than all other subjects, $F(1, 114) = 5\ 37$, $p < 022$, thus, experience had a sizable effect Finally, bus drivers performed significantly better than bartenders, $F(1, 114) = 4\ 53$, $p < 035$ As is typical for studies involving the water-level task, the error measure is associated with quite large standard deviations (ranging from 4° for male students to 18° for housewives) Thus, measures of central tendency were also computed These results are similar to those for the mean scores Median absolute error was 21 5° for waitresses, 11 5° for housewives, 2 75° for female students, 2 25° for male students, 12 0° for bartenders, and 1 0° for bus drivers Also, for the waitresses, the size of the bias grew with experience Absolute error (degrees off horizontal) and experience (years in relevant job) were correlated positively, $R = 41$, $p < 03$ (Regrettably, bartenders were not asked their number of years of experience)

During a short debriefing that took place after the experiment, among all subjects who made sizable errors, bartenders and waitresses seemed most surprised by the correct solution In fact, some of them could be convinced that the water surface remained horizontal to the ground only when the experimenter demonstrated this fact to them

Table 1 *Number of subjects who gave correct and wrong answers*

	Group						
Judgment	Waitresses	Housewives	Female students	Male students	Bartenders	Bus drivers	Total
$< -5°$	0	1	0	1	1	1	4
$> 5°$	15	13	7	3	11	4	53
$-5°$ to $5°$	5	6	13	16	8	15	63
Total	20	20	20	20	20	20	120

Note Responses were classified into three categories lines that were approximately horizontal (deviations within 5°), lines tilted more than 5° toward a parallel with the bottom of the container ($\alpha > 5°$), and lines tilted more than 5° in the opposite direction ($\alpha < -5°$)

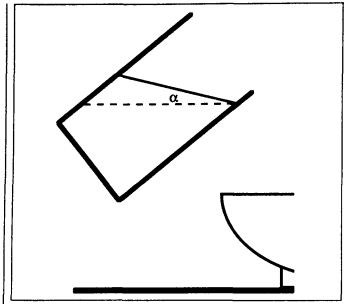

Fig 3 The same beaker as in Figure 2 The dotted line indicates the correct solution, which was produced by only one half of all subjects The solid line depicts a typical answer given by the other subjects The deviation of the answer from horizontal was measured in degrees (α)

DISCUSSION

The present studies found that waitresses and bartenders produced more errors on the water-level task than comparison groups who did not have as intensive experience with liquid-filled containers The evidence for a negative effect of profession above and beyond typical age and gender effects is clear Male bartenders performed considerably worse than male bus drivers The bus drivers were on average older (49 years) than the bartenders (33 years), thus, based on age, there was no reason to expect the former group to perform better Both groups were comparable in terms of their education, 1 of the barmen and none of the bus drivers had a college degree The fact that bartenders performed worse is a strong indicator that relevant career experience is responsible for their larger water-level bias

The comparison between housewives and waitresses was less pronounced Averaged error magnitudes were not significantly different, but the median error was almost twice as high for waitresses as for housewives Taken together, these results may be indicative of a ceiling effect in the case of the waitresses Their average error was 21 3°, which is quite dramatic compared with other studies If one looks just at those waitresses who were classified as incorrect, the average error rises to 27°, which is a little more than halfway between horizontal and parallel to the bottom of the beaker (tilted by 50°) In addition, error size and experience were correlated for the waitresses

Given that experience induces errors, what are possible explanations for this effect? How can we explain that people who frequently have horizontal liquid surfaces in front of their eyes nevertheless perform so poorly on the water-level task? Not only do they fail to profit from their experience, but they seem to be misinformed by it

Practical experience may induce a preference for the object-relative reference system that is functional in the everyday context in which that experience is acquired It is essential that bartenders and waitresses not spill drinks, thus, they have to monitor and control the discrepancy between a liquid's surface and the lip of the container and keep it within a narrow margin This context puts strong emphasis on an object-relative perspective If this perspective is predominant for expert liquid handlers, then it would seem natural that this perspective is also more likely to be employed in situations, such as the water-level task, that *require* a *different reference system* The water-level task calls for an environment-relative perspective, but the more experience people have with the other reference system, the harder it should be for them to switch over to the environment-relative system, which is not normally of functional importance to them This is, in fact, what we found

The positive correlation between experience and size of the bias suggests that it is functional for waitresses to select an object-relative perspective Hence, a good bartender or waitress is someone who is most attuned to the object-relative perspective and not to the environment-relative perspective, which is called for in the water-level task The positive correlation between years of experience and amount of error also makes a possible alternative explanation unlikely Choice of profession might correlate with some personality trait that is responsible for poor performance on the water-level task In that case, how-

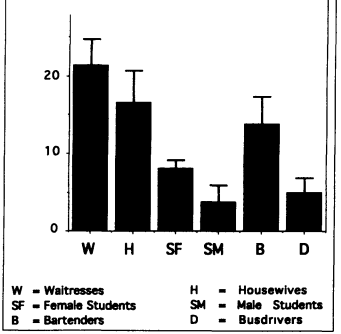

Fig 4 The average absolute error (in degrees) across all subjects by group The error bars indicate standard error of the mean

ever, all waitresses should be equally likely to make errors, which they are not Thus, the adoption of an object-relative perspective, in a context where it is functional, is likely to cause the experience effects that were found

A theory of direct perceptual attunement to dynamic invariants (Runeson & Frykholm, 1983) would have difficulties accommodating the results of the present study A more promising notion in this context is that of minimal essential information (Abernethy, 1993) Abernethy proposed that people fully process only the least amount of information that is required to carry out a desired action According to this account, waitresses and bartenders would attend only to the information that is necessary and sufficient to avoid spilling liquid while carrying glasses This information is contained in the relationship between the rim of the glass and the liquid level They would not further relate the object-relative relations to an environment-relative framework containing the structural invariant of the horizontal water level Practical experience, then, does not lead to an understanding of the physics of the situation or to attunement to a perceptual invariant Rather, practical experience consists of the least-effort acquisition of knowledge that is functional and indispensable to perform the required action In the case of the water-level problem, practical experience promotes an object-relative perspective that, in turn, evokes a frame-of-reference bias in perceiving the horizontal

Acknowledgments—We thank Susan Sugarman and one anonymous reviewer for their thoughtful comments Anette Fahle, Ina Hecht, and Isa Schinzel assisted in administering the questionnaires The research was supported by a postdoctoral grant from the Max-Planck-Institut fur Psychologische Forschung to the first author

REFERENCES

Abernethy, B (1993) Searching for the minimal essential information for skilled perception and action *Psychological Research 55* 131–138

Asch, S E , & Witkin, H A (1948) Studies in space orientations II Perception of the upright with displaced visual fields and with body tilted *Journal of Experimental Psychology 38* 455–477

Howard, I (1978) Recognition and knowledge of the water-level principle *Perception 7* 151–160

Kaiser, M K , Proffitt, D R , Whelan, S , & Hecht H (1992) Influence of animation on dynamical judgments *Journal of Experimental Psychology Human Perception and Performance 18* 669–690

Kalichman, S C (1988) Individual differences in water-level performance *Developmental Review, 8,* 273–295

Liben L S (1991) The Piagetian water-level task Looking beneath the surface In R Vasta (Ed), *Annals of Child Development* (Vol 8, pp 81–144) London Kingsley

McAfee E A & Proffitt D R (1991) Understanding the surface orientation of liquids *Cognitive Psychology 23* 483–514

Myer K A , & Hensley, J H (1984) Cognitive style, gender, and self-report of principle as predictors of adult performance on Piaget s water level task *The Journal of Genetic Psychology 144* 179–183

Piaget, J , & Inhelder B (1956) *The child s conception of space* London Routledge and Kegan Paul (Original work published 1948)

Proffitt, D R & Gilden, D L (1989) Understanding natural dynamics *Journal of Experimental Psychology Human Perception and Performance 15,* 384–393

Robert, M , & Tanguay M (1990) Perception and representation of the Euclidean coordinates in mature and elderly men and women *Experimental Aging Research, 16* 123–131

Runeson, S , & Frykholm, G (1983) Kinematic specification of dynamics as an informational basis for person-and-action perception Expectation gender recognition and deceptive intention *Journal of Experimental Psychology General 112* 585–615

(RECEIVED 4/22/94, ACCEPTED 7/7/94)