

Still More on Pulling the Goalie

ALAN WASHBURN

*Naval Postgraduate School
Monterey, California 93943-5000*

Once again the goalie is taken to task for staying in the game too long. This time dynamic programming is used to get around a questionable assumption made in previous analyses.

The rules of hockey specify that only six players can be on the ice at any one time. Normally a team will find it prudent to devote one of these players (the goalie) entirely to goal defense. The rules do not require this, however, so whichever team is behind may elect to "pull its goalie" in order to devote all six skaters to the task of scoring on the opponent. This can clearly be done too early, since the maneuver is more likely to result in a score by the opponent than by the pulling team but may nonetheless prove fruitful in the waning moments of a game that will otherwise be almost surely lost. Morrison and Wheat [1986] attribute the US victory in the 1980 Winter Olympics to employment of the tactic in the game with Sweden. The question to be examined here is "What is

the optimal time to pull the goalie?"

Background

Morrison [1976] wrote the original paper on this subject; he computed optimal pull times in a game where the opponents were evenly matched. Since 1976, three derivative papers have appeared in this journal: Morrison and Wheat [1986] in a "misapplication review" argue that Morrison [1976] mistakenly compared only the options "never pull the goalie" and "pull the goalie immediately," whereas "pull the goalie later" also ought to be considered. The "now or never" restriction causes the goalie to be pulled too early, but Morrison and Wheat correct the mistake. Erkut [1987] (after making invidious comparisons between the Edmonton Oilers and the Detroit Red Wings) generalizes the Morrison

and Wheat analysis to include the possibility that the two teams are of unequal scoring ability. In a rejoinder, Morrison calls Erkut's analysis excellent and comments ". . . in my view, he puts the finishing touch on this Poisson-based pulling-the-goalie problem." In yet another paper, Nydick and Weiss [1989] raise a statistical question about how scoring rates are to be

A team may pull its goalie to devote all six skaters to the task of scoring.

estimated for those periods when the goalie is pulled, showing that the "fixed rate" assumption and the "proportional" assumption lead to different pull times. In a comment following that paper, Morrison (perhaps somewhat wistfully) speculates that the work of Nydick and Weiss ". . . should be the final chapter on our by now beloved friend, The Goalie."

Well, here is another paper on the subject:

It's a Question of Goals

All four of the papers mentioned above adopt the objective of maximizing the probability that a particular team (hereafter the Dogs) will succeed in scoring before the other team (the Cats) and also before the game ends, the idea being that the Dogs are currently behind by one goal and have got to do something. Strictly speaking, scoring first is neither necessary nor sufficient for victory. For example, the Dogs could score first and still lose. Or the Cats could score first and then the Dogs could score three times and still win. Or the Dogs could score twice, and then the

Cats could pull their goalie and score, and so forth. Of course, all of these multiple scoring opportunities are unlikely given that the game is practically over anyway, so one might ask, "Why put up with all the additional complexity merely to include second order effects?" There are two reasons:

- (1) It isn't that hard to include second order effects.
- (2) Scoring rates are rather high when goalies are pulled (0.5 goals per minute or so), and previous authors have argued that the goalie should be pulled at up to three minutes from the end of the game. The probability that a Poisson random variable with mean $0.5 \times 3 = 1.5$ is larger than 1 is not negligible.

The second point above would be harder to make for current practice, which is to pull the goalie with roughly one minute remaining. This fact may have originally led to the decision to include only first order effects in the model, but it turns out that the first order model is self destructive in the sense that it leads to decisions for which second order effects are not negligible. Fortunately, the assumption that the two scoring processes are independent and Poisson (an assumption that will be retained), permits a dynamic programming analysis that does not require first order approximations. The Poisson assumption itself seems to be fairly solid [Mullett 1977].

Solution by Dynamic Programming (DP)

Let λ (μ) be the scoring rate for the Dogs (Cats). The memoryless property of the two independent Poisson processes means that a hockey game is characterized by state (x, t) , where x is the current relative

THE GOALIE

score ($x > 0$ means that the Dogs are ahead) and t is the time remaining (see Hlynka and Sheahan [1987]). When $x < 0$, the Dogs must decide whether (λ, μ) should be (λ_0, μ_0) or (λ_D, μ_D) , where the first pair corresponds to both goalies being present and the second pair corresponds to the Dogs' goalie being pulled. A similar choice must be made by the Cats when $x > 0$, except that the choice is between (λ_0, μ_0) and (λ_C, μ_C) . Let $F(x, t)$ be the probability that the Dogs win when both sides follow optimal policies for goalie pulling. Letting δ be a small increment of time, $F(x, t)$ must satisfy the recursive equation

$$F(x, t + \delta) = \max(\min)_{(\lambda, \mu)} \{ (\lambda\delta)F(x + 1, t) + (\mu\delta)F(x - 1, t) + (1 - \lambda\delta - \mu\delta)F(x, t) \}, \quad (1)$$

where max applies for $x < 0$, min for $x > 0$, and $(\lambda, \mu) = (\lambda_0, \mu_0)$ for $x = 0$. Given initial and boundary conditions, equation (1) can be used to calculate $F(x, \delta)$ for all x , then $F(x, 2\delta)$ for all x , and so forth, recording the optimal decision to make in each state as the computations proceed. Bellman [1957] shows that the solution of (1) converges to a function with the proper meaning as δ approaches 0. The calculations reported below utilize $\delta = 0.01$ minute.

Scoring first is neither necessary nor sufficient for victory.

$F(x, 0)$ is 0 for $x < 0$ and 1 for $x > 0$ in all forms of hockey, but the proper value for $F(0, 0)$ depends on the type of hockey being played. NHL and Olympic rules differ, and NHL rules differ between regular

season and play-offs. The base case below takes $F(0, 0)$ to be

$$F(0, 0) = [\lambda_0/(\lambda_0 + \mu_0)] [1 - \exp(-5(\lambda_0 + \mu_0))] + 0.5 \exp(-5(\lambda_0 + \mu_0)), \quad (2)$$

which corresponds to flipping a coin to decide the winner if no goal has been scored at the end of a five-minute sudden death overtime period. During the regular season, NHL rules specify that the winner gets two points and the loser gets none, except that each team gets one point in case the score is still tied at the end of a five-minute overtime. The expected number of points won is then $2F(x, t)$, so the base case corresponds tactically to NHL regular season rules in spite of the fact that no coin is actually flipped. It is also assumed in the base case that whichever side pulls its goalie increases its own scoring rate by a factor of 2.67 and the opponent's scoring rate by a factor of 7.83 [Erkut 1987].

Two excursions from the base case will be considered. In the first, $F(0, 0)$ will be taken to be $\lambda_0/(\lambda_0 + \mu_0)$, which corresponds to NHL play-off rules where the sudden death overtime period is in essence infinitely long. The second excursion retains regular season scoring rules, but takes the scoring rate of the team pulling its goalie to be 0.16/min and of the other team to be 0.47/min regardless of (λ_0, μ_0) , the "fixed rate" assumption of Nydick and Weiss [1989].

Boundary conditions for all computations reported below are that any team that succeeds in getting ahead by 10 goals at any time is an automatic winner.

Computational Results

For $(\lambda_0, \mu_0) = (0.07, 0.09)/\text{min}$ in the base case, Figure 1 shows the function $F(x, t)$ for $-3 \leq x \leq 3$ and $t \leq 10$ minutes, with the times at which the goalie (Dogs or Cats as appropriate) should be pulled being shown by hash marks on the curves.

Points worthy of note are that the superior team (Cats, in this case) that falls behind is more inclined to pull the goalie than an inferior team in the same situation, and that the further behind a team gets, the more inclined it is to pull its goalie. When behind by two goals, the Cats should pull their goalie with up to 7.7 minutes remaining. It should be understood that pulling the goalie is not irrevocable. If the down-by-two Cats pull their goalie with 7.7 min-

utes remaining and then score a goal with 7.5 minutes remaining, they should put the (rather busy) goalie right back in, pulling him again with 2.6 minutes remaining should the Cats still be one goal behind at that point. All of these events and actions are properly accounted for in the dynamic programming analysis.

Table 1 compares Erkut's times with those obtained by dynamic programming in the case where the Dogs are down by one goal. It is clear from Table 1 that there is only a slight difference between regular season and play-off pull times, and that both are substantially larger than Erkut's pull times. Even Erkut's times are larger than current practice, so the DP pull times are almost shockingly large.

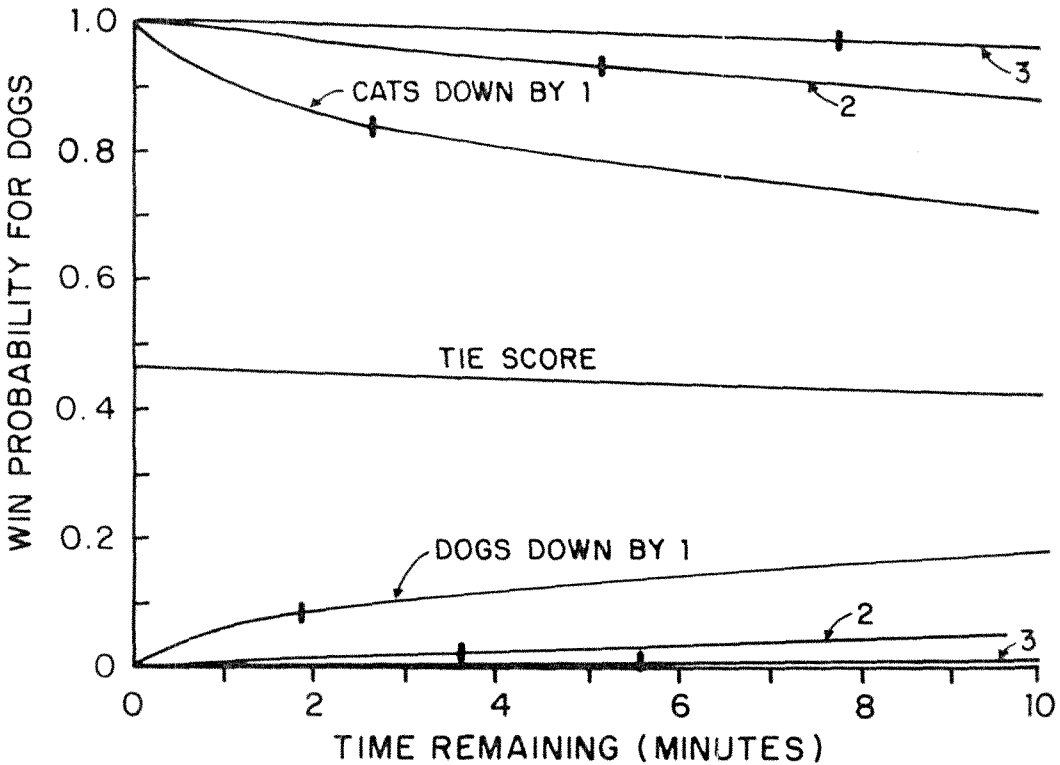


Figure 1: The probability that the Dogs win in the base case versus time remaining. Hash marks show goalie pulling times.

THE GOALIE

		μ_0		
		.05	.07	.09
λ_0	.05	2.82 (3.46, 3.46)	2.09 (2.35, 2.40)	1.66 (1.79, 1.83)
	.07	2.69 (3.72, 3.62)	2.02 (2.47, 2.47)	1.61 (1.85, 1.88)
	.09	2.58 (4.00, 3.81)	1.95 (2.59, 2.55)	1.59 (1.92, 1.92)

Table 1: Erkut's optimal goalie pulling times in minutes, with the optimal times in the base case and the play-off variation shown in parentheses.

Table 2 compares the base case with the fixed-rate variation. Whether the effect of pulling the goalie is "proportional" or "fixed rate" is evidently important, as Nydick and Weiss [1989] point out. The assumptions differ substantially. When $(\lambda_0, \mu_0) = (0.05, 0.09)/\text{min}$, for example, the scoring rates with the Dogs goalie pulled are (0.134, 0.705) in the base case or (0.16, 0.47) in the variation. The Dog/Cat scoring rate ratio jumps from 0.19 to 0.34, so the enhanced tendency of the Dogs to pull their goalie in the fixed rate variation is understandable.

And Now for a Word from Our Sponsor

Dynamic programming can evidently be used to compute goalie pull times that are optimal in the sense of maximizing the probability of winning. In this sense previously computed pull times are only approximately optimal, being based on a first order approximation. However, the increased precision does not resolve the dif-

		μ_0		
		.05	.07	.09
λ_0	.05	3.46, 3.42	2.35, 3.92	1.79, 4.59
	.07	3.72, 2.21	2.47, 2.43	1.85, 2.69
	.09	4.00, 1.46	2.59, 1.57	1.92, 1.71

Table 2: A comparison of the pull times in minutes for the "proportional" and "fixed-rate" cases.

ference between theory and practice. It exaggerates it, in fact; the DP pull times are the largest yet encountered. How to explain the difference between theory and practice?

Certainly the theory could be improved. The state vector could be augmented to include important aspects of the game other than "relative score" and "time left." The most natural of these augmentations would probably be to include an indicator of puck possession. Morrison and Wheat [1986] argue that exhaustion is also an important aspect of the game: ". . . your best players may tire if you pull the goalie when the model predicts." Perhaps some way could be found to include a measure of exhaustion in the state vector. The formulation given above is computationally trivial, so the state vector could be augmented in more than one way without running into Bellman's "curse of dimensionality."

In my humble opinion (my humility derives from the armchair nature of my acquaintance with the sport), however, none of these improvements is likely to lower the optimal pull times to the point where they agree with what currently happens in practice. We seem to have a situation where an apparently reasonable model of the closing moments of a hockey game re-

sults in recommended actions that differ strongly from conventional wisdom, while at the same time experimentation that would resolve the difference is risky, at least from the viewpoint of those who have to do the experiments. Readers who have been involved with doing or using military OR will recognize the situation. The armed forces have been confronting similar difficulties ever since OR got started in WWII, and seem to have come to the conclusion that at some point one has to tell the analysts to stop arguing and just go out and try it. Hopefully the fifth *Interfaces* paper on this subject will include a discussion of what actually happens when pull times are increased.

References

- Bellman, R. 1957, *Dynamic Programming*, Princeton University Press, Chapter 11.
- Erkut, E. 1987, "Note: More on Morrison and Wheat's 'Pulling the goalie revisited,'" *Interfaces*, Vol. 17, No. 5, pp. 121-123.
- Hlynka, M. and Sheahan, J. N. 1987, "Controlling rates in a double queue," *Naval Research Logistics*, Vol. 34, No. 4, pp. 569-577.
- Morrison, D. G. 1976, "On the optimal time to pull the goalie: A Poisson model applied to a common strategy used in ice hockey," *TIMS Studies in Management Science*, Vol. 4, North Holland, pp. 67-68.
- Morrison, D. G. and Wheat, R. D. 1986, "Misapplications reviews: Pulling the goalie revisited," *Interfaces*, Vol. 16, No. 6, pp. 28-34.
- Mullett, G. 1977, "Simeon Poisson and the National Hockey League," *The American Statistician*, Vol. 31, No. 1, pp. 8-12.
- Nydick, R. L. and Weiss, H. J. 1989, "More on Erkut's 'More on Morrison and Wheat's Pulling the goalie revisited,'" *Interfaces*, Vol. 19, No. 5, pp. 45-48.